

Tartu Ülikool  
Loodus- ja täppiseaduste valdkond  
Matemaatika ja statistika instituut

Johannes Puusepp

# EKSPEKTIIL-EKSPEKTIIL GRAAFIKUD JAOTUSE SOBIVUSE HINDAMISEKS

Matemaatilise statistika eriala

Bakalaureusetöö (9 EAP)

Juhendaja: dotsent Meelis Käärik

Tartu 2017

## **Ekspektiil-ekspektiil graafikud jaotuse sobivuse hindamiseks**

Bakalaureusetöö  
Johannes Puusepp

**Lühikokkuvõte.** Käesoleva bakalaureusetöö eesmärk on tutvustatada ekspektiile ning realiseerida nendel põhinev graafiline meetod jaotuse sobivuse hindamiseks ekspektiil-ekspektiil graafikute näol. Selleks antakse esmalt analoogia tõttu ülevaade kvantiili-dest ja kvantiil-kvantiil graafikutest. Seejärel uuritakse ekspektiile, antakse nende leidmise võrrandid levinud kahjujaotuste jaoks ning tehakse läbi diskreetse juhusliku suuruse ekspektiilide leidmise näide. Ekspektiil-ekspektiil graafikute realiseerimiseks kirjutatakse vajalik tarkvara ja antakse ülevaade selle tööpõhimõttest ning kasutamisest. Viimaks rakendatakse funktsiooni töös käsitletud teoreetiliste jaotuste sobivuse hindamiseks reaalse kahjuandmestikuga. Töö koostamisel on tuginetud Helis Puksandi magistritööle ning antud bakalaureusetööd võib pidada eelmise edasiarenduseks.

**CERCS teaduseriala:** P160 Statistika, operatsioonianalüüs, programmeerimine, finants- ja kindlustusmatemaatika

**Märksõnad:** Tõenäosusjaotused, juhuslikud suurused, R (programmeerimiskeel), ekstreemumülesanded, lähendamine, statistilised mudelid.

## **Expectile-expectile graphs for assessing the fit of a distribution**

Bachelor's thesis  
Johannes Puusepp

**Abstract.** The objective of this bachelor's thesis is to introduce expectiles and to establish a graphical method for assessing the fit of a distribution using expectiles. Because of the analogy of the statistics, the thesis begins with a review on quantiles and quantile-quantile plots. It continues by investigating expectiles and by expressing equations for finding the expectiles of common loss distributions. The expectiles of every reviewed distribution and the corresponding quantiles are visualised. The first chapter is concluded with an example of the calculation of expectiles of a discrete random variable. In order to establish expectile-expectile graphs, the necessary software is written and an overview of its working principle and operation is given. The thesis is completed by using the function to assess the fit of the reviewed theoretical distributions with the empirical distribution of a test data set.

**CERCS research specialisation:** P160 Statistics, operations research, programming, actuarial mathematics

**Keywords** Probability distributions, random numbers, R (programming language), extremum problems, approximation, statistical models.

# Sisukord

<b>Sissejuhatus</b>	<b>4</b>
<b>1 Kvantiilid, üldistatud kvantiilid ja ekspektiilid</b>	<b>5</b>
1.1 Vajalikud põhimõisted . . . . .	5
1.2 Kvantiilid ja nende kasutamine jaotuse sobivuse hindamiseks . . . . .	6
1.2.1 Kvantiili mõiste, tõlgendus ja näited . . . . .	6
1.2.2 Kvantiil-kvantiil graafiku mõiste, tõlgendus ja näide . . . . .	8
1.3 Ekspektiili mõiste, tõlgendus ja näide . . . . .	10
1.3.1 Ekspektiilid ja nende leidmine . . . . .	10
1.3.2 Ekspektiilide näiteülesanne . . . . .	22
<b>2 Jaotuse sobivuse hindamine</b>	<b>26</b>
2.1 Funktsioon EEGraaf . . . . .	26
2.1.1 Teoreetiliste ja empiiriliste ekspektiilide leidmine . . . . .	26
2.1.2 Funktsiooni rakendamine . . . . .	27
2.2 Jaotuste sobivus testandmestikuga . . . . .	30
2.2.1 Eksponentjaotus . . . . .	31
2.2.2 Log-normaaljaotus . . . . .	33
2.2.3 Pareto jaotus . . . . .	35
2.2.4 Gammajaotus . . . . .	37
2.2.5 Weibulli jaotus . . . . .	39
2.2.6 Log-gamma jaotus . . . . .	41
<b>Kokkuvõte</b>	<b>43</b>
<b>Kirjandus</b>	<b>44</b>
<b>Lisad</b>	<b>45</b>

## Sissejuhatus

Tehes otsuseid olukorras, kus tegude tagajärjed pole kindlalt teada, mõtleme tihti läbi, millised võiksid olla võimalikud tulemused ja kui tõenäoliseks me neid peame. Lihtsamatel juhtudel on võimalik tulemused ning nende esinemistõenäosused peas läbi arutada, kuid osade otsuste langetamiseks tuleb veidi rohkem vaeva näha. Näiteks juhul, kui kindlustusfirma peab otsustama, milliseid preemiaid küsida kindlustuspoliisi soetavalt kliendilt, on vaja kuidagi hinnata tõenäosust, et kliendile tuleb hüvitada poliisi kaetav kahju ning samuti tuleb prognoosida kaetava kahju suurust. Probleemi lahendamiseks otsitakse funktsiooni, mis seob juhusliku suuruse võimalike väärtustega nende esinemistõenäosused. Ühte kõnealuse funktsiooni sobivuse hindamist võimaldavat meetodit on käsitletud käesolevas töös

Arvatavasti levinuim graafiline viis jaotuse sobivuse hindamiseks on kvantiil-kvantiil graafik. Tegemist on joonisega, mis kõrvutab empiirilisi kvantiile sobitatavale jaotusele vastavate teoreetiliste kvantiilidega. Meetod on laialdast kasutust leidnud oma esmapilgul lihtsa kasutatavuse ja suure visualiseeritava informatsiooni hulga tõttu. Antud töös uurime ekspektiile ning võtame kasutusele neil põhineva graafilise meetodi jaotuse sobivuse hindamiseks. Eesmärgiks on jõuda graafikuteni, mis pööravad kvantiil-kvantiil graafikutest suuremat rõhku jaotuse ekstremaalsetele väärtustele ning võivad seetõttu kanda endas lisainformatsiooni ebasümmeetriliste jaotuste kirjeldamisel. Bakalaureuse-töö koosneb kahest suuremast peatükist.

Esimene peatükk tuletab lugejale meelde vajalikke põhimõisteid ja annab analoogia tõttu ülevaate kvantiilidest ning neil põhinevast jaotuse sobivuse hindamisest koos vastavate näidetega. Seejärel uuritakse ekspektiile ja antakse nende leidmise võrrandid levinud kahjujaotuste jaoks. Peatükis avaldatakse ka log-gamma jaotusele vastav ekspektiilide leidmise võrrand ja tehakse läbi ekspektiilide leidmise näiteülesanne diskreetse juhusliku suuruse korral.

Teises peatükis rakendatakse ekspektiile jaotuse sobivuse hindamiseks. Kuna selleks varasemad tarkvaralised lahendused puudusid, alustatakse esmalt probleemi lahendamiseks loodud tarkvara kirjeldamisest. Seejärel rakendatakse funktsiooni testandmestikule ning hinnatakse iga töös käsitletud jaotuse sobivust andmetega. Märkime, et kasutatud on reaalseid kahjukindlustusest pärinevaid kahjujuhtumeid.

Töö tugineb Helis Puksandi magistritööle (Puksand, 2015) ning selles toodud tulemusi on kasutatud graafikute realiseerimiseks vajalike teoreetiliste ekspektiilide leidmiseks. Teksti küljendamisel on kasutatud tekstitöötlusprogrammi *LaTeX* ning praktiliste ülesannete lahendamine on läbi viidud statistikatarkvara *R* abil.

Autor avaldab oma tänu töö juhendajale Meelis Käärikule asjakohaste ideede, selgete selgituste ja tähelepanelike paranduste eest.

# 1 Kvantiilid, üldistatud kvantiilid ja ekspektiilid

Esmalt toome välja töö lugemiseks vajalikud põhimõisted ning tähistused ekspektiil-ekspektiil graafikute juurde asumiseks.

## 1.1 Vajalikud põhimõisted

Käesolev alapeatükk tuletab lugejale meelde põhimõisteid ja tulemusi, mis on vajalikud edasise töö lugemisel. Kui pole viidatud teisiti, on toodud definitsioonid leitavad õpikutest (Pärna, 2013) ja (Tammeraid, 2001).

**Definitsioon 1.1.** Juhusliku suuruse  $X$  jaotusfunktsiooniks nimetatakse paremalt pidevat mittekahanevat funktsiooni  $F$ , mille korral

$$F(x) = P(X \leq x), \quad x \in \mathbb{R}.$$

Pideva juhusliku suuruse korral saame jaotusfunktsiooni tuletisena esitada tihedusfunktsiooni

$$f(x) = F'(x).$$

**Definitsioon 1.2.** Juhusliku suuruse  $X$  mediaan  $\text{Med}(X)$  on defineeritud tingimusega

$$P(X \leq \text{Med}(X)) = P(X \geq \text{Med}(X)) \geq \frac{1}{2}.$$

Juhul, kui  $X$  tihedus on sümmeetriline punkti  $a$  suhtes, siis  $EX = \text{Med}(X) = a$ .

**Definitsioon 1.3.** Olgu  $X_1$  ja  $X_2$  pidevad juhuslikud suurused tihedusfunktsioonidega  $f_{X_1}$  ja  $f_{X_2}$ . Kui kehtib võrdus

$$\lim_{x \rightarrow \infty} \frac{f_{X_1}(x)}{f_{X_2}(x)} = \infty,$$

siis öeldakse, et juhuslik suurus  $X_1$  on raskema sabaga, kui juhuslik suurus  $X_2$  (Klugman, Panjer jt, 2004: 50).

**Definitsioon 1.4.** Funktsiooni  $f : D \rightarrow \mathbb{R}$  nimetatakse alamhulgas  $Z \subset D$  rangelt kasvavaks, kui iga  $z_1, z_2 \in Z$  korral järgneb võrratusest  $z_1 < z_2$  võrratus  $f(z_1) < f(z_2)$ .

**Definitsioon 1.5.** Öeldakse, et funktsioon  $f : D \rightarrow \mathbb{R}$  on kumer intervallis  $D$ , kui iga argumendi väärtuse  $z \in D$  korral leidub  $z$  selline  $\delta$ -ümbrus, et funktsiooni  $f$  graafik on argumendi väärtustel ümbrusest  $(z - \delta, z + \delta)$  allpool funktsioonile punktis  $(z, f(z))$  tõmmatud puutujat.

**Definitsioon 1.6.** Öeldakse, et funktsioon  $f : D \rightarrow \mathbb{R}$  on tükiti lineaarne intervallis  $D$ , kui intervalli  $D$  saab jaotada lõplikuks arvuks osaintervallideks, milles  $f$  on lineaarne (Leiger, 2016:75).

## 1.2 Kvantiilid ja nende kasutamine jaotuse sobivuse hindamiseks

Järgnevas toome kvantiili kui olulise juhusliku suuruse karakteristikku mõiste, anname selle intuiitiivse tõlgenduse, toome illustreerivad näited ning tutvustame kvantiil-kvantiil graafikuid kui levinud graafilist võimalust jaotuse sobivuse hindamiseks.

### 1.2.1 Kvantiili mõiste, tõlgendus ja näited

**Definitsioon 1.7.** Juhusliku suuruse  $X$   $\alpha$ -kvantiilik, kus  $0 \leq \alpha \leq 1$ , nimetatakse sellist arvu  $q_\alpha(X)$ , mis vastab võrratusele

$$P(X < q_\alpha(X)) \leq \alpha \leq P(X \leq q_\alpha(X)).$$

Teisisõnu on kvantiil juhuslikku suurust kirjeldav statistik, millest väiksemate väärtuste esinemistõenäosus on ülimalt  $\alpha$  ning millest väiksemaid või võrdseid väärtuseid esineb vähemalt tõenäosusega  $\alpha$ .

Kasutades juhusliku suuruse jaotusfunktsiooni, saab vastavalt allikale (Koenker, 2005: 5) pidevate juhuslike suuruste puhul anda definitsiooni alternatiivsel, lühemal kujul

**Definitsioon 1.8.** Juhusliku suuruse  $X$   $\alpha$ -kvantiilik, kus  $0 \leq \alpha \leq 1$ , nimetatakse arvu  $q_\alpha(X)$ , mis on määratud võrdusega

$$q_\alpha(X) = F^{-1}(\alpha) = \inf \{x : F(x) \geq \alpha\}.$$

Definitsioonist on lihtne näha kvantiili tähendust teatava optimumina, kus minimeeritakse argumendi  $x$  väärtust selliselt, et kehtiks võrratus  $F(x) \geq \alpha$ . Juhusliku suuruse  $X$  jaoks mingis mõttes optimaalse punkti leidmiseks võime kasutada kaofunktsiooni.

Artikkel (Rosasco jt, 2003: 5) käsitleb kaofunktsiooni  $\rho$  kui funktsiooni,  $\rho : [0, \infty]$ , mis kirjeldab juhusliku suuruse  $X$  ja punkti  $u$  vahelist erinevust. Kaofunktsiooni valik toimub vastavalt lahendatavale ülesandele. Näiteks juhusliku suuruse  $X$  keskvaartuse leidmiseks sobib ruutfunktsioon,  $\rho(u) = (X - u)^2$ , mediaani leidmiseks erinevuste absoluutväärtus,  $\rho(u) = |X - u|$  ning juhul, kui meid huvitab näiteks punktide  $u$  kuumine mingisse  $\varepsilon$ -ümbrusesse, kasutatakse  $\varepsilon$ -tundlikku kaofunktsiooni, kus  $\rho(u) = \max \{|X - u| - \varepsilon, 0\}$ . Optimaalseks loetakse punkti  $u$ , mille korral  $E\rho(X - u)$  on minimaalne.

Vastavalt allikale (Bellini jt, 2013:2) saame anda kolmanda kvantiilide definitsiooni, mis vaatab kvantiile kindla kaofunktsiooni miinimumkohana ning võimaldab seeläbi mõistet hilisemalt üldistada.

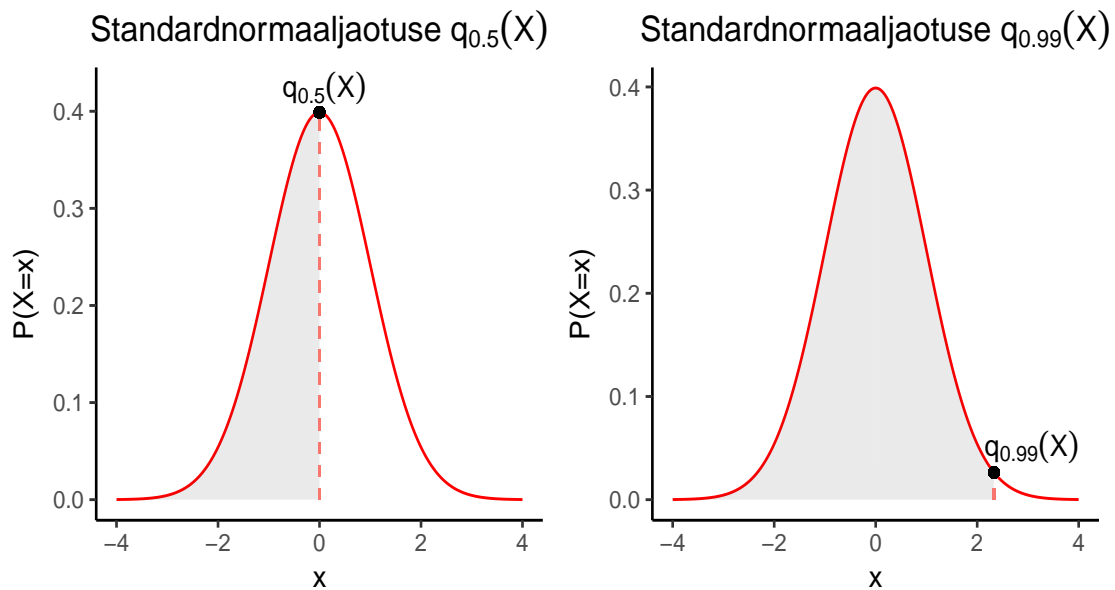
**Definitsioon 1.9.** Juhusliku suuruse  $X$   $\alpha$ -kvantiilik ( $0 \leq \alpha \leq 1$ ) nimetatakse arvu  $q_\alpha(X)$ , mis vastab tükiti lineaarse kaofunktsiooni miinimumkohale,

$$q_\alpha(X) = \arg \min_{x \in \mathbb{R}} \{ \alpha E[(X - x)^+] + (1 - \alpha) E[(X - x)^-] \},$$

kus  $x^+ := \max \{x, 0\}$  ja  $x^- := \max \{-x, 0\}$ .

Eelneva definitsiooni puhul märkame, et minimeeritavas avaldises kaalutakse hälbeid, kus  $x < X$  kaaluga  $\alpha$  ja selliseid, kus  $x > X$  kaaluga  $1 - \alpha$ . See vastab kvantiili intuiitvsele tähendusele. Näiteks võttes  $\alpha = 1 - \alpha = 0,5$ , korrutame mõlemasuunalisi hälbeid sama kordajaga ehk kaalume neid võrdselt. Avaldist minimeerides saame sellise  $x \in \mathbb{R}$  väärtuse, millest keskmiselt pooled juhusliku suuruse realisatsioonid on suuremad ja pooled väiksemad. Paneme tähele, et eelnev ühtib juhusliku suuruse mediaani definitsiooniga, ja seega  $q_{0,5}(X) = \text{Med}(X)$ .

**Näide 1.10.** Vaatame kvantiilide leidmise näidet, võttes juhuslikuks suuruseks standardse normaaljaotusega juhusliku suuruse,  $X \sim N(0, 1)$ . Leiame antud juhusliku suuruse 0,5-kvantiili  $q_{0,5}(X)$  ja 0,99-kvantiili  $q_{0,99}(X)$ , kasutades definitsiooni 1.8. Standardse normaaljaotuse tabelist saame  $\alpha = 0,5$  korral  $\inf \{x : F(x) \geq 0,5\} = 0$  ning seega  $q_{0,5}(X) = 0$ . Paneme tähele, et tegemist on nullpunkti suhtes sümmeetrilise jaotusega ning tulemus vastab definitsioonile 1.2. Fikseerides  $\alpha = 0,99$  näeme, et  $\inf \{x : F(x) \geq 0,99\} = 2,33$  ja seega  $q_{0,99}(X) = 2,33$ . Teisisõnu on 99% antud juhusliku suuruse realisatsioonidest väiksemad arvust 2,33.



Joonis 1: Standardse normaaljaotuse kvantiilid

Selgituseks toome joonise, millel on kvantiilid tähistatud katkendjoonega normaaljaotuse tihedusfunktsiooni graafikul. Tumedamalt on märgitud kvantiilist väiksemad väärtused.

**Näide 1.11.** Vaatame kvantiilide leidmise näidet, kus juhuslikuks suuruseks on viie järjestikuse mündiviske tulemusena saadud kirjade koguarv. Tegemist on binoomjaotusega juhusliku suurusega,  $X \sim \text{Bin}(n = 5, p = 0,5)$ , mille jaotuse võime anda ka järgneva tabelina:

Kirjade arv ( $x_i$ )	0	1	2	3	4	5
Tõenäosusfunktsiooni väärtus, $P(X = x_i)$	$\frac{1}{32}$	$\frac{5}{32}$	$\frac{5}{16}$	$\frac{5}{16}$	$\frac{5}{32}$	$\frac{1}{32}$
Kumulatiivne tõenäosus, $P(X \leq x_i)$	$\frac{1}{32}$	$\frac{3}{16}$	$\frac{1}{2}$	$\frac{13}{16}$	$\frac{31}{32}$	1

Tabel 1: Järjestikuste mündiviske jaotus

Leiame kvantiilid, mis vastavad  $\alpha$  väärtustele 0,5 ja 0,9, kasutades definitsiooni 1.7. Selleks fikseerime esmalt  $\alpha = 0,5$  ja näeme, et võrratus  $P(X < x_1) \leq 0,5 \leq P(X \leq x_1)$  kehtib iga  $x_1 \in (2, 3]$  korral. Seega sobib mündiviske 0,5-kvantiiliks iga poollõiku  $(2, 3]$  kuuluv arv. Järgnevalt määrame  $\alpha = 0,9$  ja märkame, et  $P(X < x_2) \leq 0,9 \leq P(X \leq x_2)$  kehtib iga  $x_2 \in (3, 4]$  korral ning seega sobib mündiviske 0,9-kvantiiliks iga poollõiku  $(3, 4]$  kuuluv arv.

### 1.2.2 Kvantiil-kvantiil graafiku mõiste, tõlgendus ja näide

Antud alapeatükis vaatleme kvantiil-kvantiil graafikut, mis on tuntud graafiline vahend empiiriliste andmete jaotuste omavaheliseks võrdlemiseks või andmete jaotuse kohta käivate hüpoteeside testimiseks. Eesmärgipärasuse huvides vaatame edaspidi olukordi, kus soovime hinnata mõne teoreetilise jaotuse sobivust andmete empiirilise jaotusega.

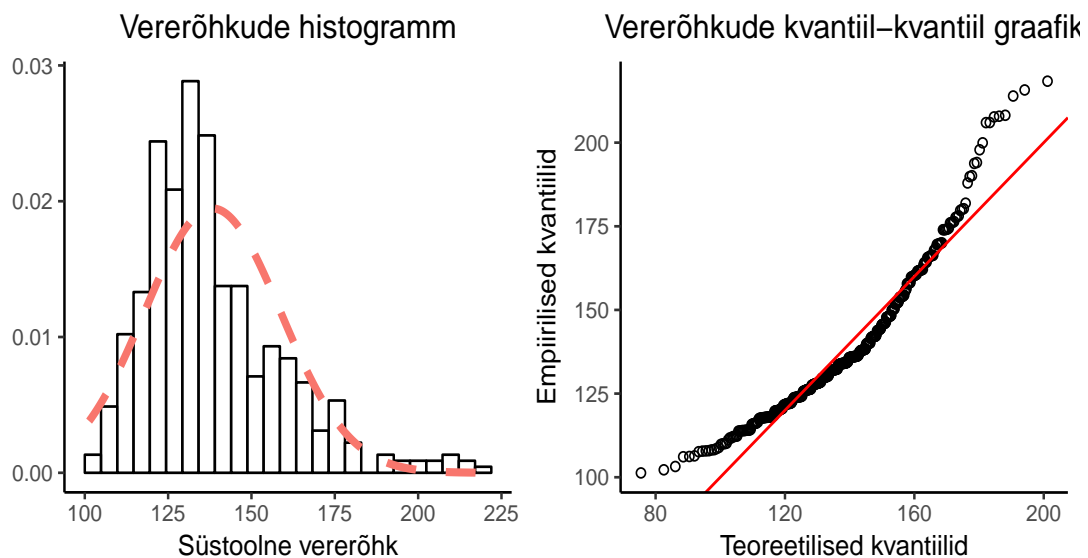
Kvantiil-kvantiil graafikuks nimetatakse joonist, mis saadakse kõrvutades andmete empiirilisi kvantiile hüpoteetilise jaotuse teoreetiliste kvantiilidega. Tegemist on graafikuga, mille  $y$ -teljel on uuritava tunnuse järjestatud väärtused ja  $x$ -teljel nendele vastavad ühtlaselt jaotatud teoreetilised kvantiilid. Olukorras, kus teoreetiline jaotus langeb kokku andmete jaotusega, asuvad kvantiilidele vastavad punktid nurgapoolitajal ehk teisisõnu empiirilised kvantiilid langevad kokku teoreetiliste kvantiilidega.

Selleks, et tagada teoreetiliste kvantiilide ühtlane jaotatus (täheenduses, et iga kahe teoreetilise kvantiili vahele jääb ühtlane tõenäosusmass), lähtutakse vaatlusele  $x_i$  vastava teoreetilise kvantiili  $q_i$  graafikule märkimisel järgnevast, allikast (Gusnanto, 2011) leitatavast valemist:

$$q_i = F^{-1} \left( \frac{i - 0,5}{n} \right).$$



**Näide 1.12.** Kontrollime kvantiil-kvantiil graafiku abil hüpoteesi andmete pärinemisest normaaljaotusest. Selleks kasutame näiteandmestikuna 462 inimese mõõdetud süstoolse vererõhu andmeid (Hastie jt, 2009). Kasutatava andmestiku keskväärus on  $EX = 138,4$  ja standardhälve  $\sqrt{DX} = 20,5$ . Kuna normaaljaotusega juhusliku suuruse  $X \sim N(\mu, \sigma^2)$  korral  $EX = \mu$  ja  $DX = \sigma^2$ , on hüpoteesiks andmete pärinemine jaotusest  $N(138,4, 20,5^2)$ . Andmete empiirilist jaotust visualiseerime ka histogrammi abil, millele on punase joonega märgitud hüpoteesjaotuse tihedusfunktsiooni graafik ning mille tulpadealune pindala on võrdne ühega. Kvantiil-kvantiil graafikul on sirge, millel paiknemise korral empiirilised kvantiilid võrduksid teoreetiliste kvantiilidega, märgitud punase joonega ning joonised on antud kõrvuti.



Joonis 2: Vererõhkude histogramm ning kvantiil-kvantiil graafik

Näeme, et andmed ei sobi siiski normaaljaotusega. Kvantiil-kvantiil graafikul ootaksime teoreetilise jaotuse andmetega sobivuse korral teoreetiliste kvantiilide sarnaseid väärtuseid empiiriliste statistikute. Teisisõnu peaksid punktid normaaljaotuse sobivuse korral järgima telgedevahelist punast sirget. Kuna antud juhul punktid eristuvad selgelt telgedevahelisest sirgest, valimikvantiilid on teoreetilistest väiksemad keskmistel argumentide  $\alpha$  väärtustel ning suuremad ülejäänud osas, saame ümber lükata andmete sobivuse normaaljaotusega. Histogramm kinnitab kvantiil-kvantiil graafiku põhjal antud hinnangut. Nii tulpdiagrammi parempoolsete tulpade tihedusfunktsiooni joonest kõrgemal paiknemisest kui empiiriliste kvantiilide suurematest väärtustest võrreldes teoreetiliste kvantiilidega kõrgetel  $\alpha$  väärtustel näeme, et empiirilistel andmetel on sobitatud normaaljaotusega võrreldes raskem saba.

## 1.3 Ekspektiili mõiste, tõlgendus ja näide

Antud alapeatükis anname esmalt üldistatud kvantiilide ja ekspektiilide definitsioonid, sõnastame ning tõestame ühe ekspektiilide tähtsa omaduse ja kirjeldame ekspektiilide leidmiseks vajalikku protsessi. Seejärel tutvustame levinud kahjujaotuseid ja anname neile vastavad ekspektiilide leidmise vörrandid. Iga jaotuse kohta esitame graafiku, millel on toodud mõnele parameetrite kombinatsioonile vastavad ekspektiilid ja kvantiilid ning kirjeldame, kuidas parameetrite manipuleerimine mõjutab statistikute muutumist. Samuti selgitame, kuidas on statistikute lõikepunkti abil võimalik hinnata jaotuse saba raskust. Alapeatüki lõpetab ekspektiilide leidmise näide diskreetse juhusliku suuruse korral. Kui pole viidatud teisiti, on lähtutud allikast (Bellini jt, 2013).

### 1.3.1 Ekspektiilid ja nende leidmine

Defineerime üldistatud kvantiili mõiste, mis on teadaolevalt esmakordselt kasutusele võetud alapeatüki algallikas (Bellini jt, 2013) ning omab töös tähtsat kohta, sest on nii kvantiilide kui ekspektiilide üldistus. Teisisõnu on töös vaadeldavad juhuslike suuruste karakteristikud järgneva mõiste erijuhud.

**Definitsioon 1.13.** Juhusliku suuruse  $X$   $\alpha$ -üldistatud kvantiiliks ( $0 \leq \alpha \leq 1$ ) nimetatakse arvu  $x_\alpha(X)$ , mis on defineeritud järgneva minimeerimisülesande lahendina:

$$x_\alpha(X) = \arg \min_{x \in \mathbb{R}} \left\{ \alpha E[\phi_1((X - x)^+)] + (1 - \alpha) E[\phi_2((X - x)^-)] \right\},$$

kus  $\phi_1$  ja  $\phi_2$  on kumerad rangelt kasvavad funktsioonid, mille korral  $\phi_i(0) = 0$  ja  $\phi_i(1) = 1$  iga  $i \in \{1, 2\}$  korral. Märkime, et mitme miinimumkoha korral nimetatakse üldistatud kvantiilideks kõiki minimeerimisülesande lahendeid.

Ekspektiilid on üldistatud kvantiilide erijuht, mille korral on funktsioonid  $\phi_i$ ,  $i \in \{1, 2\}$  ruutfunktsioonid.

**Definitsioon 1.14.** Juhusliku suuruse  $X$   $\alpha$ -ekspektiiliks ( $0 \leq \alpha \leq 1$ ) nimetatakse arvu  $e_\alpha(X)$ , mis on defineeritud järgmiselt:

$$e_\alpha(X) = \arg \min_{x \in \mathbb{R}} \left\{ \alpha E[(X - x)^+]^2 + (1 - \alpha) E[(X - x)^-]^2 \right\}.$$

Toome omaduse, mis seob juhusliku suuruse ekspektiilid selle keskvärtusega.

**Lause 1.15.** Juhusliku suuruse  $X$  0,5-ekspektiil  $e_{0,5}(X)$  on võrdne juhusliku suuruse keskvärtusega,  $e_{0,5}(X) = EX$

**Tõestus.** Kirjutame välja ekspektilide definitsiooni  $\alpha = 0,5$  korral:

$$e_{0,5}(X) = \arg \min_{x \in \mathbb{R}} \left\{ 0,5E[(X-x)^+]^2 + 0,5E[(X-x)^-]^2 \right\}.$$

Kuna vastavalt definitsioonile iga  $x \in \mathbb{R}$  korral  $((X-x)^+)^2 = 0$  või  $((X-x)^-)^2 = 0$  ning  $((X-x)^+)^2 + ((X-x)^-)^2 = (X-x)^2$ , saame avaldist lihtsustada:

$$e_{0,5}(X) = \arg \min_{x \in \mathbb{R}} \{0,5E[(X-x)^2]\}. \quad (1)$$

Diferentseerime minimeeritavat funktsiooni ja võrdsustame selle nulliga:

$$E(X-x) = 0 \Rightarrow E(X) = E(x) = x.$$

Oleme näidanud, et  $e_{0,5}(X) = EX$ .

Märgime, et toodud omadus on põhjuseks, miks ekspektilid võiksid mängida olulist rolli ebasümmeetriliste jaotuste kirjeldamisel. Nimelt selgitab tõestatud omadus ekspektilide võimet arvestada ebasümmeetriliste jaotuste ekstremaalsete väärtustega, mis jääksid mediaani ümber paiknevate kvantiilide puhul tähelepanuta.

Vormistame lausena tulemuse, mis vaatab ekspektili võrrandi lahenditena. Lauset rakendatakse hilisemalt log-gamma jaotusele vastava ekspektilide leidmise võrrandi avaldamisel.

**Lause 1.16.** Juhusliku suuruse  $X$   $\alpha$ -ekspektilideks on kõik võrrandi

$$\alpha E[(X-x)^+] + (\alpha-1) E[(X-x)^-] = 0$$

lahendid. Tulemuseks on üldistatud kvantiilide juhul leitav allikas (Bellini jt, 2013:6-7)

Järgnevalt defineerime mõned levinud kahujaotused, anname nende keskväärtuste ja dispersioonide analüütilised kujud ning esitame toodud jaotustele vastavad ekspektilide võrrandid. Eksponentjaotuse, log-normaaljaotuse, Pareto jaotuse, gammajaotuse ja Weibulli jaotuse jaoks on võrrandite leidmine läbi tehtud magistritöös (Puksand, 2015: 24-36) ning log-gamma jaotuse ekspektilide leidmiseks vajaliku võrrandi avaldamine on käesoleva töö autori panus. Kõigile ekspektilide leidmiseks kasutatavatele võrranditele järgneb joonis, millel on visualiseeritud uuritava jaotuse ekspektilide muutumist mõne parameetrite kombinatsiooni korral. Võrdlusmomendi tekitamiseks on graafikutel toodud ka vastavad kvantiilid. Ekspektilid on tähistatud katkend- ja kvantiilid pideva joonega ning parameetrite väärtused on märgitud legendil. Argumendi  $\alpha$  väärtused kuuluvad eksponentjaotuse joonisel lõiku  $[0; 0,99]$  ja kõigi muude vaadeldud jaotuste graafikutel lõiku  $[0; 0,95]$ . Kasutatud jaotuste definitsioonid ja olulisemad karakteristikud on leitavad allikatest (Pärna, 2013) ja (Hogg ja Klugman, 1984).

**Definitsioon 1.17.** Öeldakse, et juhuslik suurus  $X$  on eksponentjaotusega,  $X \sim \text{Exp}(\lambda)$ , kui tema tihedusfunktsioon  $f_X(x)$  on kujul

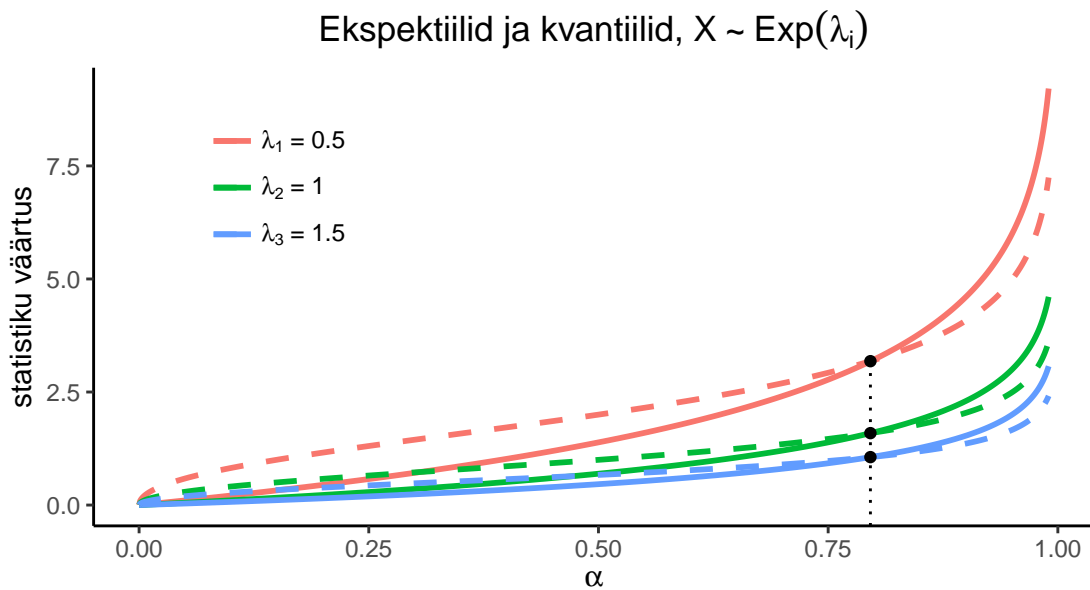
$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{kui } x \geq 0 \\ 0, & \text{kui } x < 0. \end{cases} \quad (2)$$

Sellisel juhul avaldub juhusliku suuruse  $X$  keskvärtus kujul  $EX = \frac{1}{\lambda}$  ning tema dispersioon  $DX = \frac{1}{\lambda^2}$ .

**Lause 1.18.** Olgu juhuslik suurus  $X$  eksponentjaotusega,  $X \sim \text{Exp}(\lambda)$  ehk olgu tema tihedusfunktsioon kujul (2). Sellisel juhul rahuldavad  $X$  ekspektiilid võrrandit

$$(2\alpha - 1) \frac{e^{-\lambda e_\alpha(X)}}{\lambda} + (\alpha - 1) \left( e_\alpha(X) - \frac{1}{\lambda} \right) = 0.$$

Joonisel on toodud eksponentjaotusega juhusliku suuruse  $X \sim \text{Exp}(\lambda)$  ekspektiilid ja kvantiilid parameetri  $\lambda$  kolme väärtuse korral.



Joonis 3: Eksponentjaotuse ekspektiilid ja kvantiilid

Näeme, et statistikute väärtused on jaotuse parameetriga pöördvõrdelises seoses. Kvantile ja ekspektiile võrreldes märkame, et ekspektiilide väärtused ületavad argumenti  $\alpha$

väiksemate väärtuste korral kvantiilide väärtuseid, kusjuures joonte tõusunurgad erinevad enim just argumenti määramispiirkonna alguses. Võrreldes statistikuid  $\alpha = 0,5$  korral, näeme, et antud parameetrite korral ületab keskvärtus märgatavalt mediaani. Kvantiilide ja ekspektiilide graafikute lõikekohaks on iga valitud  $\lambda$  korral  $\alpha = 0,8$ . Intuiitivseks selgituseks ekspektiilide suuremale väärtusele madalatel  $\alpha$  väärtustel on jaotuse tihedusfunktsiooni ebasümmeetrilisus. Tihedus langeb eksponentsiaalselt argumenti  $x$  kasvades, kuid on positiivne kogu  $x$ -telje mittenegatiivses osas. Kuna kvantiile mõjutab vaid statistikust väiksemate realisatsioonide osakaal, arvestatakse harvaesinevate väga suurte väärtustega madala argumenti  $\alpha$  korral vaid proportsionaalselt nende esinemissagedusega. Ekspektiilid võtavad arvesse lisaks ekstremaalsete realisatsioonide osakaalule ka nende väärtuseid ning see võib osutada soositavaks omaduseks näiteks kindlustuskahjude kirjeldamisel.

**Definitsioon 1.19.** Olgu juhuslik suurus  $Y$  normaaljaotusega,  $Y \sim N(\mu, \sigma^2)$  ja  $X = e^Y$ . Sellisel juhul öeldakse, et  $X$  on log-normaaljaotusega,  $X \sim LN(\mu, \sigma^2)$ , kus  $\mu$  ja  $\sigma^2$  on  $Y$ -le vastavad parameetrid ning tihedusfunktsioon  $f_X(x)$  on kujul

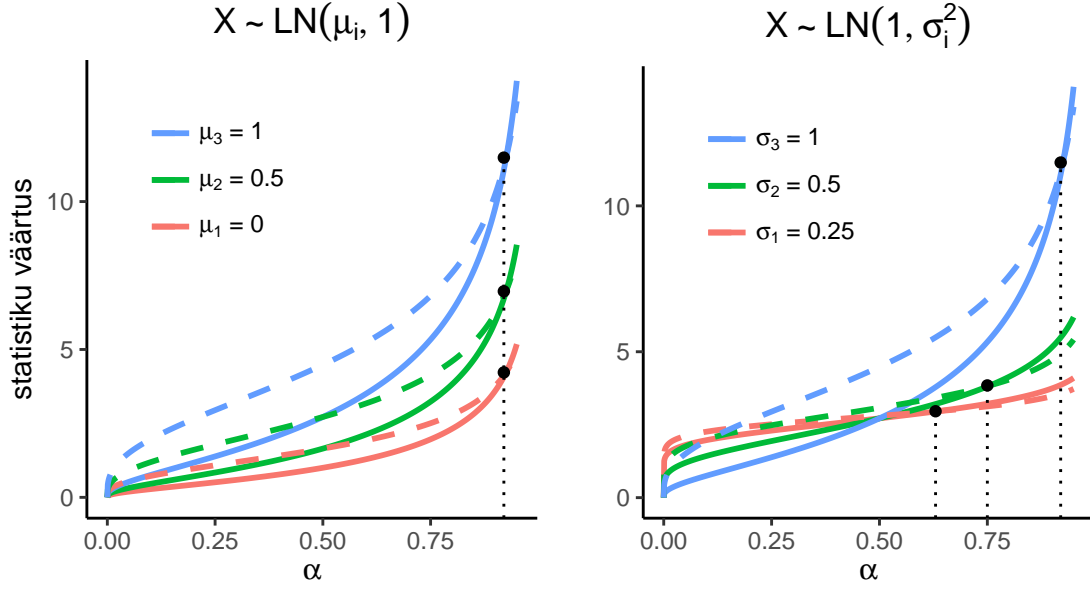
$$f_X(x) = \begin{cases} \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}, & \text{kui } x > 0 \\ 0, & \text{kui } x \leq 0. \end{cases} \quad (3)$$

Log-normaaljaotusega juhusliku suuruse  $X$  keskvärtus avaldub kujul  $EX = e^{\mu + \frac{\sigma^2}{2}}$  ja dispersioon  $DX = e^{2(\mu + \sigma^2)}(e^{\sigma^2} - 1)$ .

**Lause 1.20.** Olgu juhuslik suurus  $X$  log-normaaljaotusega,  $X \sim LN(\mu, \sigma^2)$  ehk olgu tema tihedusfunktsioon kujul (3). Sellisel juhul rahuldavad  $X$  ekspektiilid võrrandit

$$(1 - 2\alpha)e^{\mu + \frac{\sigma^2}{2}} \phi\left(\frac{\ln e_\alpha(X) - \sigma^2 - \mu}{\sigma}\right) + (2\alpha - 1)e_\alpha(X) \phi\left(\frac{\ln e_\alpha(X) - \mu}{\sigma}\right) + \alpha\left(e^{\mu + \frac{\sigma^2}{2}} - e_\alpha(X)\right) = 0.$$

Graafik visualiseerib log-normaaljaotusega juhusliku suuruse  $X \sim LN(\mu, \sigma^2)$  ekspektiilie ja kvantiile. Vasakpoolsel joonisel on muudetud parameetri  $\mu$  väärtust ning parempoolsel joonisel parameetrit  $\sigma^2$ .



Joonis 4: Log-normaaljaotuse ekspektiilid ja kvantiilid

Jooniselt näeme, et log-normaaljaotuse korral on statistikud jaotuse esimese parameetriga  $\mu$  võrdelises seoses, kuid hajuvust määrav teine parameeter  $\sigma^2$  mõjutab statistikuid erisugustel  $\alpha$  väärtustel erinevalt. Paneme tähele, et jaotuse keskvväärtus suureneb oluliselt hajuvusparameetri suurematel väärtustel, kuid mediaan on võrdne kõigi  $\sigma^2$  väärtuste korral. Lisaks näeme, et ekspektiilide ja kvantiilide löikekoha määrab jaotuse teine parameeter. Löikekoha paiknemine suurematel  $\alpha$  väärtustel võiks vihjata jaotuse saba raskusele. Tähelepanu tuleks pöörata ka statistikute muutumisele argumenti määramispiirkonna alguses. Kuna jaotuse tihedusfunktsioon on log-normaaljaotusel nullilähedastel argumenti  $x$  väärtustel kasvav, on ka joonte tõusunurgad madalatel argumenti väärtustel oluliselt suuremad eksponentjaotusest. Teisisõnu on väga madalate log-normaaljaotusega juhusliku suuruse realisatsioonide esinemine vähem ootuspärane. Märkime, et omadus võib osutuda tähtsaks teises peatükis läbiviidaval jaotuse sobivuse hindamisel.

**Definitsioon 1.21.** Ütleme, et juhuslik suurus  $X$  on Pareto jaotusega,  $X \sim Pa(\gamma, \beta)$ , kus  $\gamma, \beta > 0$ , kui tema tihedusfunktsioon  $f_X(x)$  on kujul

$$f_X(x) = \begin{cases} \frac{\gamma\beta^\gamma}{(\beta+x)^{\gamma+1}}, & \text{kui } x \geq 0 \\ 0, & \text{kui } x < 0. \end{cases} \quad (4)$$

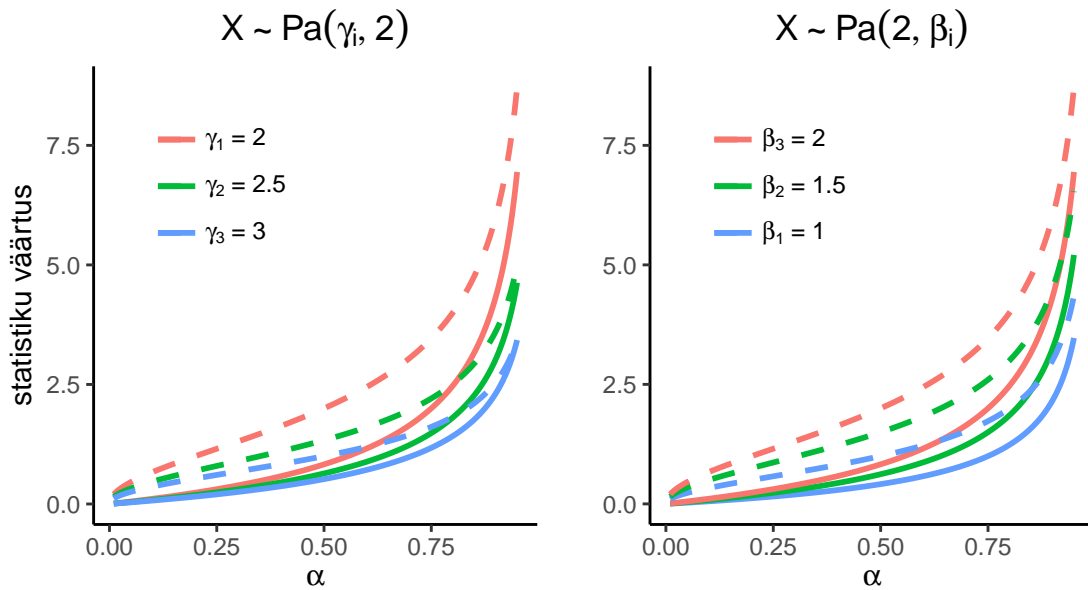
Pareto jaotusega juhusliku suuruse  $X$  keskvärtus avaldub kujul  $EX = \frac{\beta}{\gamma - 1}$ , kui  $\gamma >$

1 ja dispersioon  $DX = \frac{\gamma\beta^2}{(\gamma - 1)^2(\gamma - 2)}$ , kui  $\gamma > 2$ .

**Lause 1.22.** Olgu juhuslik suurus  $X$  Pareto jaotusega,  $X \sim Pa(\gamma, \beta)$  ehk teisisõnu olgu tema tihedusfunktsioon kujul (4). Sellisel juhul rahuldavad tema ekspektiilid võrrandit

$$(2\alpha - 1) \frac{(\beta + e_\alpha(X))^{-\gamma}}{\gamma - 1} + (1 - 2\alpha) e_\alpha(X) \beta^\gamma (\beta + e_\alpha(X))^{-\gamma} + (\alpha - 1) \left( e_\alpha(X) - \frac{\beta}{\gamma - 1} \right) = 0.$$

Järgnevalt vaatame ekspektiilide ja kvantiilide muutumist Pareto jaotusega juhusliku suuruse  $X \sim Pa(\gamma, \beta)$  korral. Vasakpoolne graafik vaatab statistikuid erinevate  $\gamma$  ja parempoolne erinevate  $\beta$  väärtuste korral.



Joonis 5: Pareto jaotuse ekspektiilid ja kvantiilid

Jooniselt on näha, et Pareto jaotusega juhusliku suuruse korral on jaotuse esimene parameeter statistikutega pöördvõrdelises ning teine võrdelises seoses. Paneme tähele, et toodud argumendi  $\alpha$  väärtustel puudub statistikutel lõikepunkt. See annab alust arvata, et tegemist on eksponentjaotusest ja log-normaaljaotusest raskema sabaga jaotusega.

**Definitsioon 1.23.** Öeldakse, et juhuslik suurus  $X$  on gammajaotusega,  $X \sim \Gamma(\gamma, \beta)$ , kus  $\gamma > 0$  on kujuparameeter ja  $\beta = \frac{1}{\theta} > 0$  on pöördskaalaparameeter, kui tema tihedusfunktsioon  $f_X(x)$  on kujul

$$f_X(x) = \begin{cases} \frac{\beta^\gamma}{\Gamma(\gamma)} x^{\gamma-1} e^{-\beta x}, & \text{kui } x > 0 \\ 0, & \text{kui } x \leq 0, \end{cases} \quad (5)$$

kus  $\Gamma(\gamma)$  on gammafunktsioon, mis on määratud valemiga:

$$\Gamma(\gamma) = \int_0^\infty t^{\gamma-1} e^{-t} dt.$$

Gammajaotuse keskväärtus avaldub kujul  $EX = \frac{\gamma}{\beta}$  ning dispersioon  $DX = \frac{\gamma}{\beta^2}$ .

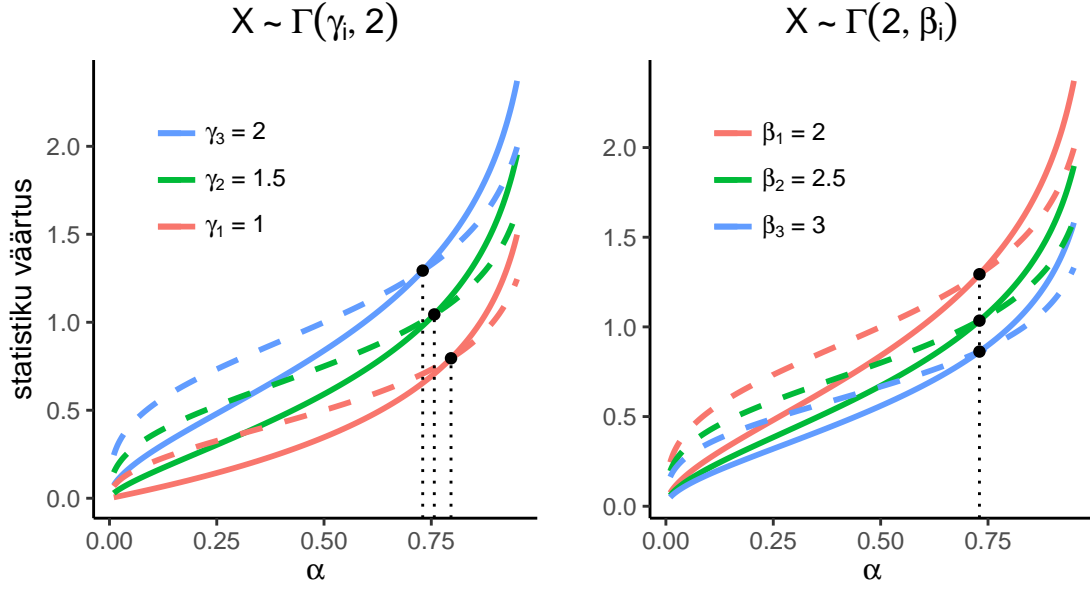
Märgime, et naturaalarvulise parameetri  $\gamma$  puhul on gammajaotusega juhuslik suurus vaadeldav  $\gamma$  sõltumatu eksponentjaotusega (parameetriga  $\lambda = \beta$ ) juhusliku suuruse summana (Pishro-Nik, 2014: 192).

**Lause 1.24.** Olgu juhuslik suurus  $X$  gammajaotusega,  $X \sim \Gamma(\gamma, \beta)$  ehk olgu tema tihedusfunktsioon kujul (5). Sellisel juhul rahuldavad juhusliku suuruse  $X$  ekspektiilid võrrandit

$$\begin{aligned} (1 - 2\alpha) \frac{e_\alpha(X) \Gamma(\gamma, \beta e_\alpha(X))}{\Gamma(\gamma)} + (2\alpha - 1) \frac{\Gamma(\gamma + 1, \beta e_\alpha(X))}{\beta \Gamma(\gamma)} \\ + (\alpha - 1) \left( e_\alpha(X) - \frac{\gamma}{\beta} \right) = 0. \end{aligned}$$

Toome joonise, millele on kantud gammajaotusega juhusliku suuruse  $X \sim \Gamma(\gamma, \beta)$  ekspektiilid ja kvantiilid parameetrite  $\gamma$  ja  $\beta$  erinevate kombinatsioonide korral.





Joonis 6: Gammajaotuse ekspektiilid ja kvantiilid

Graafikult näeme, et statistikud on jaotuse esimese parameetriga  $\gamma$  võrdelises ja teise parameetriga  $\beta$  pöördvõrdelises seoses. Kvantiilide ja ekspektiilide lõikepunkti mõjutab parameeter  $\gamma$ , mille suurematel väärtustel lõikuvad statistikud juba madalamate argumentide  $\alpha$  väärtuste korral.

**Definitsioon 1.25.** Öeldakse, et juhuslik suurus  $X$  on Weibulli jaotusega,  $X \sim We(k, \lambda)$ , kus  $k > 0$  on kujuparameeter ja  $\lambda > 0$  on skaalaparameeter, kui tema tihedusfunktsioon  $f_X(x)$  on kujul

$$f_X(x) = \begin{cases} \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k}, & \text{kui } x \geq 0 \\ 0, & \text{kui } x < 0. \end{cases} \quad (6)$$

Sel juhul on juhusliku suuruse  $X$  keskvärtus  $EX = \lambda \Gamma\left(1 + \frac{1}{k}\right)$  ja dispersioon

$$DX = \lambda^2 \left[ \Gamma\left(1 + \frac{2}{k}\right) - \left(\Gamma\left(1 + \frac{1}{k}\right)\right)^2 \right].$$

Märgime, et

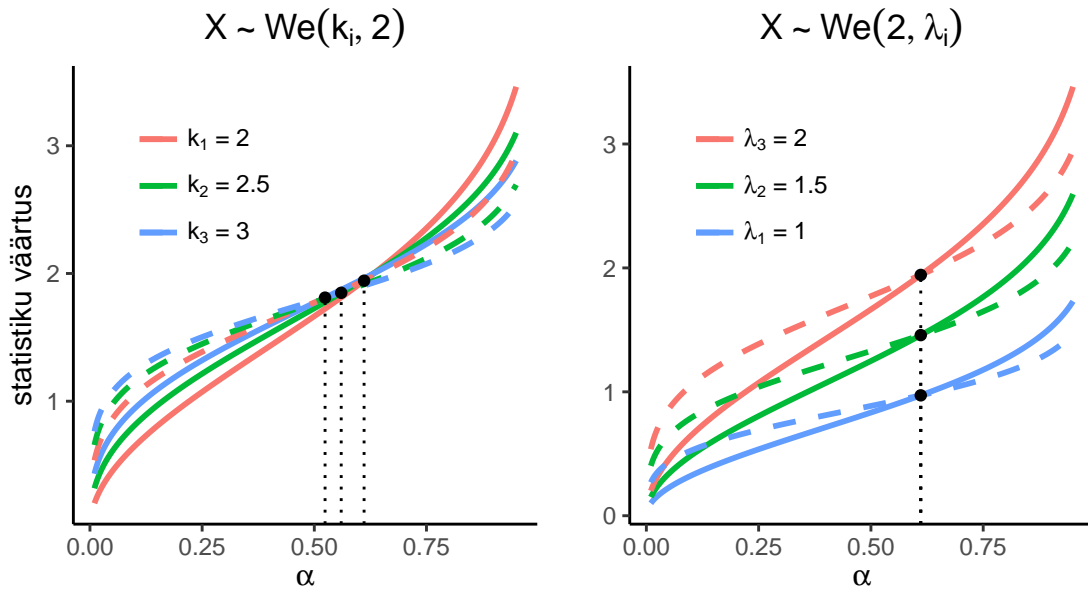
- juhul  $k < 1$  on Weibulli jaotuse tihedusfunktsiooni graafik eksponentjaotuse ja Pareto jaotuse tihedusfunktsioonide graafikute vahel;

- juhul  $k > 1$  on Weibulli jaotuse tihedusfunktsioonil kergem saba kui eksponentjaotusel;
- juhul  $k = 1$  langeb Weibulli jaotus kokku eksponentjaotusega (Kaasik ja Käärrik, 2015: 34).

**Lause 1.26.** Olgu juhuslik suurus  $X$  Weibulli jaotusega ehk olgu tema tihedusfunktsioon kujul (6). Sellisel juhul rahuldavad juhusliku suuruse  $X$  ekspektiilid võrrandit

$$(2\alpha - 1)\lambda\Gamma\left(1 + \frac{1}{k}, \left(\frac{e_\alpha(X)}{\lambda}\right)^k\right) + (1 - 2\alpha)e_\alpha(X)e^{-\left(\frac{e_\alpha(X)}{\lambda}\right)^k} + (\alpha - 1)\left(e_\alpha(X) - \lambda\Gamma\left(1 + \frac{1}{k}\right)\right) = 0.$$

Kanname joonisele Weibulli jaotusega juhusliku suuruse  $X$ ,  $X \sim We(k, \lambda)$  ekspektiilid ja kvantiilid parameetrite  $k$  ja  $\lambda$  erinevate väärtuste korral.



Joonis 7: Weibulli jaotuse ekspektiilid ja kvantiilid

Graafikult paistab, et esimese parameetri  $k$  suuremad väärtused kasvatavad statistikuid väiksemate  $\alpha$  väärtuste korral ning vähendavad suuremate  $\alpha$  väärtuste puhul. Samuti

määrab esimene parameeter ekspektilide ja kvantiilide lõikekohad, mis on suuremate parameetri  $k$  väärtuste korral argumendi  $\alpha$  väiksemate väärtuste juures. Jaotuse teine parameeter  $\lambda$  on statistikutega võrdelises seoses ning ekspektilide ja kvantiilide lõikekohta ei mõjuta.

Järgnevalt avaldame ekspektilide leidmise võrrandi log-gamma jaotuse jaoks. Selleks defineerime esmalt huvipakkuva kahjujaotuse.

**Definitsioon 1.27.** Olgu juhuslik suurus  $Y$  gammajaotusega,  $Y \sim \Gamma(\gamma, \beta)$  ja  $X = e^Y$ . Sellisel juhul öeldakse, et  $X$  on log-gamma jaotusega,  $X \sim LG(\gamma, \beta)$ , kus  $\gamma > 0$  on  $Y$ -le vastav kujuparameeter,  $\beta = \frac{1}{\theta} > 0$  on  $Y$ -le vastav pöördskaalaparameeter ning tihedusfunktsioon  $f_X(x)$  on kujul

$$f_X(x) = \begin{cases} \frac{\gamma^\beta (\ln x)^{\beta-1}}{\Gamma(\beta) x^{\gamma+1}}, & \text{kui } x \geq 1 \\ 0, & \text{kui } x < 1. \end{cases} \quad (7)$$

Log-gamma jaotusega juhusliku suuruse  $X$  keskvaartus  $EX = \left(\frac{\gamma}{\gamma-1}\right)^\beta$ , kui  $\gamma > 1$  ja dispersioon  $DX = \left(\frac{\gamma}{\gamma-2}\right)^\beta - \left(\frac{\gamma}{\gamma-1}\right)^{2\beta}$ , kui  $\gamma > 2$ .

Ekspektilide võrrandi avaldamiseks kasutame lauset 1.16, mille kohaselt rahuldavad ekspektilid võrrandit

$$\alpha \int_{e_\alpha(X)}^\infty (t - e_\alpha(X)) \frac{\gamma^\beta (\ln t)^{\beta-1}}{\Gamma(\beta) t^{\gamma+1}} dt + (\alpha - 1) \int_0^{e_\alpha(X)} (e_\alpha(X) - t) \frac{\gamma^\beta (\ln t)^{\beta-1}}{\Gamma(\beta) t^{\gamma+1}} dt = 0. \quad (8)$$

Avaldame esmalt võrrandi (8) esimese integraali

$$\begin{aligned} & \alpha \int_{e_\alpha(X)}^\infty (t - e_\alpha(X)) \frac{\gamma^\beta (\ln t)^{\beta-1}}{\Gamma(\beta) t^{\gamma+1}} dt \\ &= \frac{\alpha \gamma^\beta}{\Gamma(\beta)} \left( \int_{e_\alpha(X)}^\infty \frac{(\ln t)^{\beta-1}}{t^\gamma} dt - e_\alpha(X) \int_{e_\alpha(X)}^\infty \frac{(\ln t)^{\beta-1}}{t^{\gamma+1}} dt \right), \end{aligned}$$

milles omakorda leiame esmalt esimese integraali. Selleks paneme tähele, et muutujavahetusega  $u = \ln(t) \Rightarrow \frac{du}{dt} = \frac{1}{t}$  saame  $t^{1-\gamma} = e^{(1-\gamma)u} \Rightarrow (\ln t)^{\beta-1} = u^{\beta-1}$  ja seega

$$\int_{e_\alpha(X)}^{\infty} \frac{(\ln t)^{\beta-1}}{t^\gamma} dt = \int_{\ln(e_\alpha(X))}^{\infty} u^{\beta-1} e^{(1-\gamma)u} du.$$

Avaldamaks integraali gammafunktsioonina, kasutame allikast (Gradshteyn ja Ryzhik, 2007: 340) pärinevat avaldist

$$\int_u^{\infty} x^n e^{-\mu x} dx = \mu^{-n-1} \tilde{\Gamma}(n+1, \mu u), \quad (9)$$

kus  $\tilde{\Gamma}(\gamma, x)$  on ülemine mittetäielik gammafunktsioon,  $\tilde{\Gamma}(\gamma, x) = \int_x^{\infty} t^{\gamma-1} e^{-t} dt$  ja saame

$$\int_{\ln(e_\alpha(X))}^{\infty} u^{\beta-1} e^{(1-\gamma)u} du = \frac{\tilde{\Gamma}(\beta, (\gamma-1) \ln(e_\alpha(X)))}{(\gamma-1)^\beta}.$$

Järgmisena leiame uuritava integraali teise liidetava. Selleks teeme samuti esmalt muutujavahetuse  $u = \ln(t)$ :

$$\int_{e_\alpha(X)}^{\infty} \frac{(\ln t)^{\beta-1}}{t^{\gamma+1}} dt = \int_{\ln(e_\alpha(X))}^{\infty} u^{\beta-1} e^{-\gamma u} du.$$

Kasutades avaldist (9), saame integraaliks

$$\int_{\ln(e_\alpha(X))}^{\infty} u^{\beta-1} e^{-\gamma u} du = \frac{\tilde{\Gamma}(\beta, \gamma \ln(e_\alpha(X)))}{\gamma^\beta}.$$

Kokkuvõttes saame võrrandi (8) esimeseks integraaliks

$$\begin{aligned} & \frac{\alpha \gamma^\beta}{\Gamma(\beta)} \left( \int_{e_\alpha(X)}^{\infty} \frac{(\ln t)^{\beta-1}}{t^\gamma} dt - e_\alpha(X) \int_{e_\alpha(X)}^{\infty} \frac{(\ln t)^{\beta-1}}{t^{\gamma+1}} dt \right) \\ &= \frac{\alpha \gamma^\beta}{\Gamma(\beta)} \left( \frac{\tilde{\Gamma}(\beta, (\gamma-1) \ln(e_\alpha(X)))}{(\gamma-1)^\beta} - e_\alpha(X) \frac{\tilde{\Gamma}(\beta, \gamma \ln(e_\alpha(X)))}{\gamma^\beta} \right). \end{aligned}$$

Avaldame nüüd võrrandi (8) teise integraali

$$\begin{aligned}
& (\alpha - 1) \int_0^{e_\alpha(X)} (e_\alpha(X) - t) \frac{\gamma^\beta (\ln t)^{\beta-1}}{\Gamma(\beta) t^{\gamma+1}} dt \\
&= \frac{(\alpha - 1) \gamma^\beta}{\Gamma(\beta)} \int_0^{e_\alpha(X)} (e_\alpha(X) - t) \frac{(\ln(t))^{\beta-1}}{t^{\gamma+1}} dt \\
&= \frac{(\alpha - 1) \gamma^\beta}{\Gamma(\beta)} \left( e_\alpha(X) \int_0^{e_\alpha(X)} \frac{(\ln t)^{\beta-1}}{t^{\gamma+1}} dt - \int_0^{e_\alpha(X)} \frac{(\ln t)^{\beta-1}}{t^\gamma} dt \right),
\end{aligned}$$

mille puhul toetume allikast (Gradshteyn ja Ryzhik, 2007: 340) leitavale avaldisele

$$\int_0^u x^n e^{-\mu x} dx = \mu^{-n-1} \gamma(n+1, \mu u),$$

kus  $\gamma(\gamma, x)$  on alumine mittetäielik gammafunktsioon,  $\gamma(\gamma, x) = \int_0^x t^{\gamma-1} e^{-t} dt$ .  
Avaldame liidetavad integraalid kasutades taaskord muutujavahetust  $u = \ln(t)$ :

$$\begin{aligned}
\int_0^{e_\alpha(X)} \frac{(\ln t)^{\beta-1}}{t^{\gamma+1}} dt &= \int_0^{\ln(e_\alpha(X))} u^{\beta-1} e^{-\gamma u} du = \frac{\gamma(\beta, \gamma \ln(e_\alpha(X)))}{\gamma^\beta} \\
\int_0^{e_\alpha(X)} \frac{(\ln t)^{\beta-1}}{t^\gamma} dt &= \int_0^{\ln(e_\alpha(X))} u^{\beta-1} e^{(1-\gamma)u} du = \frac{\gamma(\beta, (\gamma-1) \ln(e_\alpha(X)))}{(\gamma-1)^\beta}.
\end{aligned}$$

Ning kirjutame välja võrrandi (8) teise integraali:

$$\begin{aligned}
& \frac{(\alpha - 1) \gamma^\beta}{\Gamma(\beta)} \left( e_\alpha(X) \int_0^{e_\alpha(X)} \frac{(\ln t)^{\beta-1}}{t^{\gamma+1}} dt - \int_0^{e_\alpha(X)} \frac{(\ln t)^{\beta-1}}{t^\gamma} dt \right) \\
&= \frac{(\alpha - 1) \gamma^\beta}{\Gamma(\beta)} \left( e_\alpha(X) \frac{\gamma(\beta, \gamma \ln(e_\alpha(X)))}{\gamma^\beta} - \frac{\gamma(\beta, (\gamma-1) \ln(e_\alpha(X)))}{(\gamma-1)^\beta} \right).
\end{aligned}$$

Kokkuvõttes saame võrrandi (8) esitada kujul

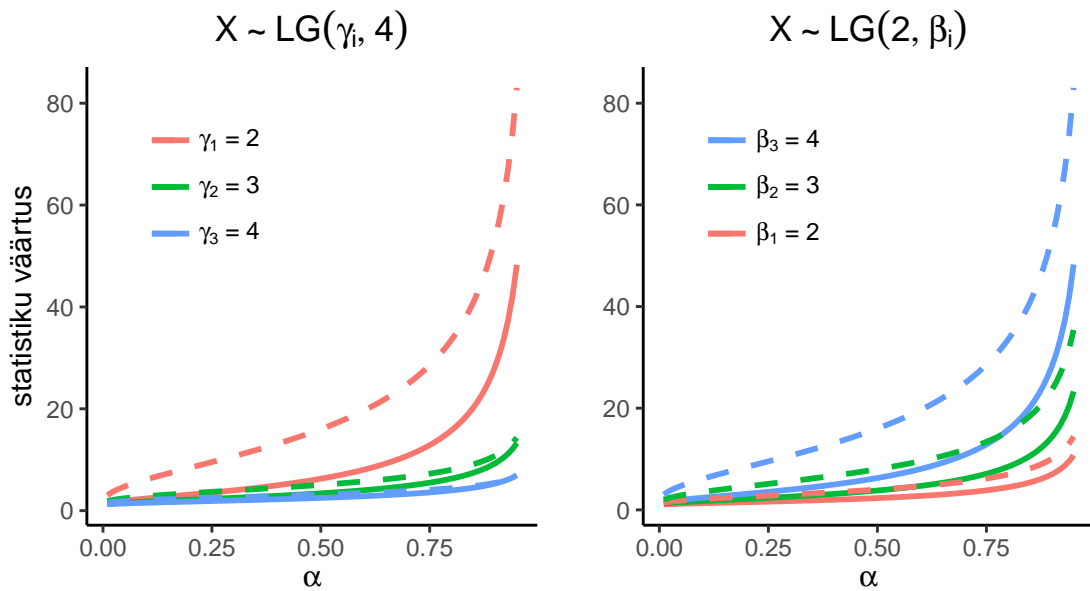
$$\begin{aligned}
& \frac{\alpha \gamma^\beta}{\Gamma(\beta)} \left( \frac{\tilde{\Gamma}(\beta, (\gamma-1) \ln(e_\alpha(X)))}{(\gamma-1)^\beta} - e_\alpha(X) \frac{\tilde{\Gamma}(\beta, \gamma \ln(e_\alpha(X)))}{\gamma^\beta} \right) \\
&+ \frac{(\alpha - 1) \gamma^\beta}{\Gamma(\beta)} \left( e_\alpha(X) \frac{\gamma(\beta, \gamma \ln(e_\alpha(X)))}{\gamma^\beta} - \frac{\gamma(\beta, (\gamma-1) \ln(e_\alpha(X)))}{(\gamma-1)^\beta} \right) = 0.
\end{aligned}$$

Vormistame saadud tulemuse lausena.

**Lause 1.28.** Olgu juhuslik suurus  $X$  log-gamma jaotusega,  $X \sim LG(\gamma, \beta)$  ehk olgu tema tihedusfunktsioon kujul (7). Sel juhul rahuldavad juhusliku suuruse  $X$  ekspektilid võrrandit

$$\frac{\alpha\gamma^\beta}{\Gamma(\beta)} \left( \frac{\tilde{\Gamma}(\beta, (\gamma-1)\ln(e_\alpha(X)))}{(\gamma-1)^\beta} - e_\alpha(X) \frac{\tilde{\Gamma}(\beta, \gamma\ln(e_\alpha(X)))}{\gamma^\beta} \right) + \frac{(\alpha-1)\gamma^\beta}{\Gamma(\beta)} \left( e_\alpha(X) \frac{\gamma(\beta, \gamma\ln(e_\alpha(X)))}{\gamma^\beta} - \frac{\gamma(\beta, (\gamma-1)\ln(e_\alpha(X)))}{(\gamma-1)^\beta} \right) = 0.$$

Visualiseerime saadud võrrandi lahenditena leitud log-gamma jaotuse ekspektilid ja neile vastavaid kvantiile joonisega.



Joonis 8: Log-gamma jaotuse ekspektilid ja kvantiilid

Graafikult näeme, et jaotuse esimene parameeter  $\gamma$  on negatiivses seoses statistikute väärtusega ning teine parameeter  $\beta$  positiivses. Ekspektilidel ja kvantiilidel puudub joonisel toodud  $\alpha$  väärtustel lõikekoht ning see annab alust arvata, et tegemist on raske sabaga jaotusega.

### 1.3.2 Ekspektilide näiteülesanne

Selles alapeatükis teeme läbi ekspektilide leidmise diskreetse juhusliku suuruse korral.

**Näide 1.29.** Vaatame ekspektiilide leidmise näidet, kasutades taaskord juhusliku suurusena viie järjestikuse mündiviske tulemusena saadud kirjade arvu, mille jaotus on antud näites 1.11. Leiame ekspektiilid mis vastavad  $\alpha$  väärtustele 0,5 ja 0,9. Selleks fikseerime esmalt  $\alpha = 0,5$  ja kasutame võrrandit (1):

$$\begin{aligned} e_{0,5}(X) &= \arg \min_{x \in \mathbb{R}} \{0,5E[(X-x)^2]\} \\ &= \arg \min_{x \in \mathbb{R}} \left\{ \frac{1}{32}(0-x)^2 + \frac{5}{32}(1-x)^2 \right. \\ &\quad \left. + \frac{5}{16}(2-x)^2 + \frac{5}{16}(3-x)^2 + \frac{5}{32}(4-x)^2 + \frac{1}{32}(5-x)^2 \right\} \\ &= \arg \min_{x \in \mathbb{R}} \{x^2 - 5x + 7,5\}. \end{aligned}$$

Diferentseerime sulgudes olevat polünoomi ning miinimumkoha leidmiseks võrdsustame tuletise nulliga:

$$2x - 5 = 0.$$

Saame, et  $x = 2,5$  ja seega  $e_{0,5}(X) = 2,5$ . Vastavalt lausele 1.15 peab kehtima  $E(X) = 2,5$ . Teades, et  $X \sim \text{Bin}(n = 5, p = 0,5)$ , kontrollime tulemust:

$$E(X) = np = 2,5.$$

Järgnevalt fikseerime  $\alpha = 0,9$  ja kasutame definitsiooni 1.14:

$$\begin{aligned} e_{0,9}(X) &= \arg \min_{x \in \mathbb{R}} \left\{ 0,9E[(X-x)^+]^2 + 0,1E[(X-x)^-]^2 \right\} \\ &= \arg \min_{x \in \mathbb{R}} \left\{ 0,9 \left( \frac{1}{32}(\max(0-x, 0))^2 + \frac{5}{32}(\max(1-x, 0))^2 \right. \right. \\ &\quad \left. \left. + \frac{5}{16}(\max(2-x, 0))^2 + \frac{5}{16}(\max(3-x, 0))^2 + \frac{5}{32}(\max(4-x, 0))^2 \right. \right. \\ &\quad \left. \left. + \frac{1}{32}(\max(5-x, 0))^2 \right) + 0,1 \left( \frac{1}{32}(\max(x-0, 0))^2 + \frac{5}{32}(\max(x-1, 0))^2 \right. \right. \\ &\quad \left. \left. + \frac{5}{16}(\max(x-2, 0))^2 + \frac{5}{16}(\max(x-3, 0))^2 + \frac{5}{32}(\max(x-4, 0))^2 \right. \right. \\ &\quad \left. \left. + \frac{1}{32}(\max(x-5, 0))^2 \right) \right\}. \end{aligned}$$

Leidmaks miinimumkohta, vaatame esmalt juhtu, kus  $x = 5$ . Siis on esimene, kaaluga  $\alpha = 0,9$  liidetav 0 ning teine liidetav  $0,1 \cdot \left( \frac{25}{32} + \frac{80}{32} + \frac{45}{16} + \frac{20}{16} + \frac{5}{32} + 0 \right) = 7,5$ . Juhul, kui  $x \in [4, 5)$ , saame minimeeritavaks avaldiseks

$$0,9 \cdot \frac{1}{32} (5-x)^2 + 0,1 \left( \frac{1}{32} (x-0)^2 + \frac{5}{32} (x-1)^2 + \frac{5}{16} (x-2)^2 + \frac{5}{16} (x-3)^2 + \frac{5}{32} (x-4)^2 + \frac{1}{32} (0)^2 \right) = 0,125x^2 - 0,75x + 1,375.$$

Diferentseerime avaldist ning võrdsustame selle nulliga. Saame

$$0,25x - 0,75 = 0,$$

millest  $x = 3$ . Kuna uuritava ruutfunktsiooni ekstreemum ei asu poollõiguses  $[4, 5)$ , siis on funktsioon antud intervallis monotoonne ja selle poollõigu miinimum saavutatakse kohal  $x = 4$ .

Edasi vaatame juhtu, kui  $x \in [3, 4)$ . Minimeeritavaks avaldiseks on

$$0,9 \left( \frac{1}{32} (5-x)^2 + \frac{5}{32} (4-x)^2 \right) + 0,1 \left( \frac{1}{32} (x-0)^2 + \frac{5}{32} (x-1)^2 + \frac{5}{16} (x-2)^2 + \frac{5}{16} (x-3)^2 \right) = 0,25x^2 - 1,75x + 3,375.$$

Minimeerime ja saame  $x = 3,5$  ning sellisel juhul on algse avaldise väärtuseks  $0,3125 < 7,5$ . Veendumaks ekspektiili väärtuses leiame minimeeritava avaldise väärtused ka ülejäänud poollõikudes.

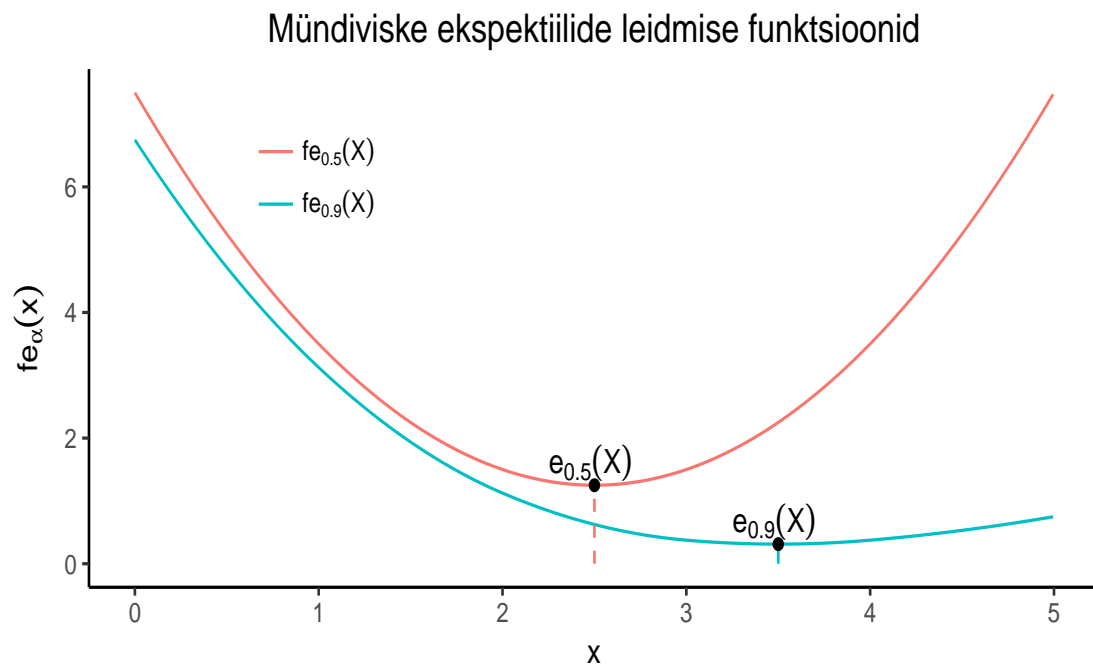
Olgu  $x \in [2, 3)$ . Sel juhul saame minimeeritavaks avaldiseks  $0,5x^2 - 3,25x + 5,625$  ning selle miinimumkohaks  $x = 3,25$ . Kuna leitud ekstreemum ei kuulu vaadeldavas poollõiku, on ekspektiilide leidmise funktsioon antud poollõiguses monotoonne ja selle poollõigu miinimum asub kohal  $x = 3$ .

Vaatame olukorda, kus  $x \in [1, 2)$ . Siis on minimeeritav avaldis  $0,75x^2 - 4,25x + 6,625$ , mille miinimumkoht on  $x = 2\frac{5}{6}$ . Saadud väärtus ei kuulu jällegi vaadeldavas poollõiku.

Viimaks kontrollime juhtu, kus  $x \in [0, 1)$ . Sel juhul on minimeeritavaks avaldiseks  $0,875x^2 - 4,5x + 6,75$  ja selle miinimumkohaks on  $2\frac{4}{7} \notin [0, 1)$ .

Leidsime 0,9-ekspektiili väärtuse, milleks on  $e_{0,9}(X) = 3,5$ . Illustreerime tulemusi joonisega:





Joonis 9: Mündiviske ekspektiilide leidmise funktsioonid

Jooniselt näeme kahte parabooli, mille miinimumkohtadeks on otsitavate ekspektiilide väärtused. Märgime, et eri poollõikudes on tegemist erinevate ruutfunktsioonide graafikutega.

## 2 Jaotuse sobivuse hindamine

Peatükis hindame käsitletud jaotuste sobivust testandmestikuga. Kuna ekspektiil-ekspektiil graafikute realiseerimiseks teadaolevad varasemad rakendused puuduvad, alustame selleks loodud funktsiooni kirjeldamisest. Seejärel rakendame loodud rakendust testandmestikule ning kasutame seda kõigi töös vaadeldud jaotuste sobivuse hindamiseks.

### 2.1 Funktsioon **EEGraaf**

Alapeatükis kirjeldame loodud funktsiooni, mis võimaldab hinnata kasutaja valitud teoreetilise jaotuse sobivust andmestikuga ekspektiilide ning kvantiilide abil. Tarkvara annab võimaluse luua nelja tüüpi graafikuid:

- ekspektiil-ekspektiil graafik;
- kvantiil-kvantiil graafik;
- normeeritud histogramm koos sobitatava jaotuse tihedusfunktsiooniga;
- teoreetiliste ja empiiriliste ekspektiilide ning kvantiilide muutumist kirjeldav graafik.

Ekspektiil-ekspektiil ning kvantiil-kvantiil graafikutel on teoreetiliste ja empiiriliste statistikute võrdlemise lihtsustamiseks telgedest võrdsel kaugusel asuvad koordinaadid tähistatud punase sirgjoonega. Teisisõnu paiknevad mõlema joonise korral teoreetilise ja valimijaotuse kokkulangevuse korral punktid punasel telgedevahelisel sirgel ning olulise hälbumise korral antud sirgjoonest saame väita, et teoreetiline jaotus ei sobi andmetega. Viimasel joonisetüübil, mis võrdleb omavahel teoreetilisi ja empiirilisi statistikuid, on ekspektiilid tähistatud katkend- ja kvantiilid pideva joonega ning teoreetilised statistikud on märgitud punase ja empiirilised näitajad rohelise joonevärviga. Funktsioon on realiseeritud, kasutades tarkvarapaketti *R*. Autorile teadaolevalt on ekspektiilide leidmiseks vajalik rakendus varasemalt loodud vaadeldavatest jaotustest eksponentjaotuse, log-normaaljaotuse ja gammajaotuse jaoks.

#### 2.1.1 Teoreetiliste ja empiiriliste ekspektiilide leidmine

Teoreetilised ekspektiilid on leitud lausetes 1.18-1.28 toodud võrrandite lahenditena. Võrrandite lahendamiseks on võetud nende vasakust poolest absoluutväärtus ning tulemust minimeeritud. Miinimumkohale vastav argumendi väärtus on uuritava võrrandi lahendiks. Empiiriliste ekspektiilide arvutamisel on lähtutud definitsioonis 1.14 toodud minimeerimisülesandest.

Avaldiste miinimumkohtade leidmiseks on kasutatud Austraalia matemaatik Richard Peirce Brent'i poolt loodud numbrilist meetodit, mille põhjalik kirjeldus on leitav allikast (Brent, 1973: 72-79). Algoritm minimeerib funktsiooni  $f : D \rightarrow \mathbb{R}$  kasutaja poolt määratud lõigus  $[a; b]$ , vajamata sealjuures informatsiooni minimeeritava funktsiooni tuletiste kohta. Kuna üldjuhul ei hinnata funktsiooni lõigu otspunktides, on vajalik vaid vahemiku  $(a; b)$  kuulumine funktsiooni määramispiirkonda  $D$ . Meetodi koondumine on garanteeritud ning selleks ei tule funktsiooni hinnata rohkem kui

$$\frac{2}{\log_2 [(1 + \sqrt{5})]} \left[ \log_2 \left( \frac{b - a}{tol} \right) \right]^2$$

korral, kus  $tol$  on kasutaja soovitud täpsus.

Kuna kõigi uuritavate kahjujaotuste tihedusfunktsioon on positiivne vaid kindlas piirkonnas, saame seada ka jaotusele vastavad itereerimispiirkonna alguspunktid  $a$ . Otsupunkti  $b$  valikul on ekspektiile lähendatud uuritavale jaotusele, selle parameetritele ja argumendile  $\alpha$  vastava kvantiiliga, mida on korrutatud empiirilistel katsetustel leitud konstandiga ning millele on liidetud sobiv vabaliige. Sobitatava jaotuse parameetrite hindamiseks on kasutatud suurima tõepära meetodit.

### 2.1.2 Funktsiooni rakendamine

Järgnevalt kirjeldame loodud rakenduse kasutamist, selgitame kasutaja määratavate parameetrite tähendust ja nende vaikeväärtusi ning toome kaks näidet funktsiooni väljast.

Funktsiooni `EEGraaf` kasutamisel on kasutajal vajalik väärtustada kaks parameetrit: `jaotus` ja `andmestik`. Lisaks toodud kohustuslikele parameetritele on kasutajal võimalik rakenduse tööd mõjutada järgnevate argumentide abil: `hist`, `kvant`, `ekspKvant`, `esimene`, `cToke`, `cParam`, `cSkaala`, `tArv`, `aLimEksp`, `aLimKvant`, `aLimHist`.

Selgitame funktsiooni argumentide tähendusi ja anname nende vaikeväärtused:

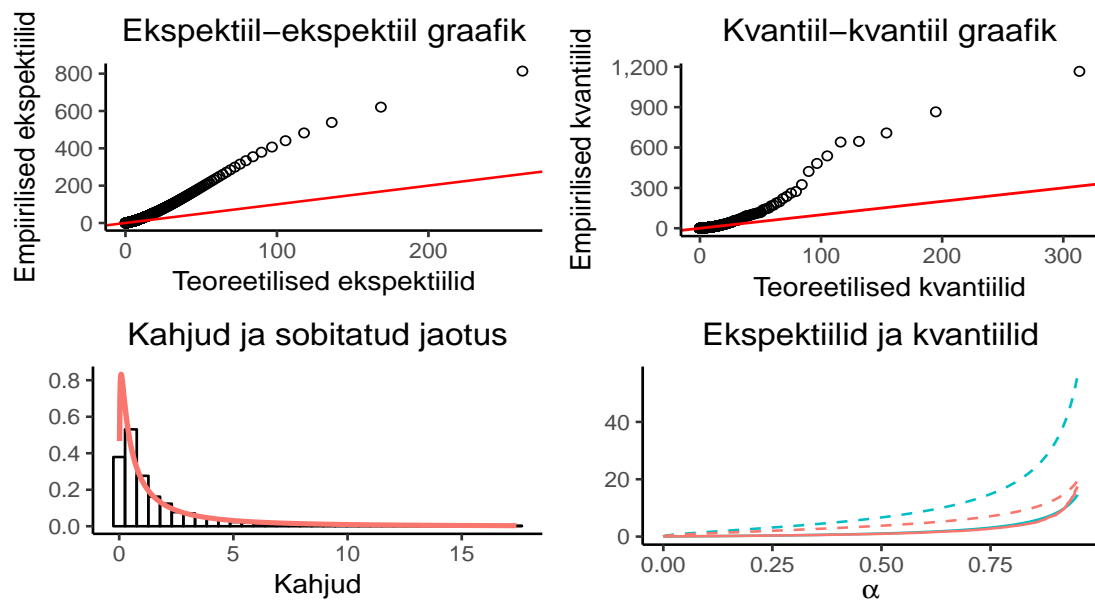
- `jaotus` - sõne, mille lubatavad väärtused on "eksponent", "log-norm", "Pareto", "gamma", "Weibull" ja "log-gamma";
- `andmestik` - reaalarvudest koosnev vektor klassist *numeric*;
- `hist` - binaarne tunnus, mis määrab histogrammi loomise. Võimalikud väärtused on `TRUE` ja `FALSE`, vaikeväärtus `TRUE`;
- `kvant` - binaarne tunnus, mis määrab kvantiil-kvantiil graafiku loomise. Võimalikud väärtused on `TRUE` ja `FALSE`, vaikeväärtus `TRUE`;

- `ekspKvant` - binaarne tunnus, mis määrab teoreetiliste ja empiiriliste ekspektilide-kvantiilide graafiku loomise. Võimalikud väärtused on `TRUE` ja `FALSE`, vaikeväärtus `TRUE`;
- `esimene` - binaarne tunnus, mis määrab, kas empiirilised ekspektilid leitakse esimest korda. Võimaldab vältida korduvaid arvutusi funktsiooni järjestikustel rakendamistel ühele andmestikule. Võimalikud väärtused on `TRUE` ja `FALSE`, vaikeväärtus `TRUE`;
- `cToke` - arv, mida kasutatakse ekspektili ülalt tõkestava kvantiili skaleerimisel, vaikeväärtus 1000;
- `cParam` - arv, millega viiakse läbi skaalateisendus parameetrite hindamisel, vaikeväärtus 10000;
- `cSkaala` - arv, mis lubab teisendada skaalat loodavatel graafikutel, vaikeväärtus 1000.
- `tArv` - naturaalarv, mis määrab histogrammi tulpade arvu, vaikeväärtus 35;
- `aLimEksp` - ülemine tõke  $\alpha$  väärtustele ekspektil-ekspektil graafikutel. Vaikeväärtus 1.
- `aLimKvant` - ülemine tõke  $\alpha$  väärtustele kvantiil-kvantiil graafikutel. Vaikeväärtus 1.
- `aLimHist` - ülemine tõke  $\alpha$  väärtustele histogrammi ning teoreetiliste ja empiiriliste ekspektilide-kvantiilide graafiku  $x$ -teljel. Vaikeväärtus 0,95

Illustreerimaks funktsiooni töö tulemust, alustame rakenduse väljakutsest, millel on väärtustatud vaid kaks esimest parameetrit. Märgime, et edasistel joonistel on parameetri `cSkaala` väärtuseks vaikeväärtus 1000 ning seetõttu on jooniste telgedel märgitud kahjude suurused antud tuhandetes ühikutes.

```
EEGraaf("log-norm", andmed)
```

Toodud viisil funktsiooni rakendamise tulemuseks saame neljast joonisest koosneva graafiku:



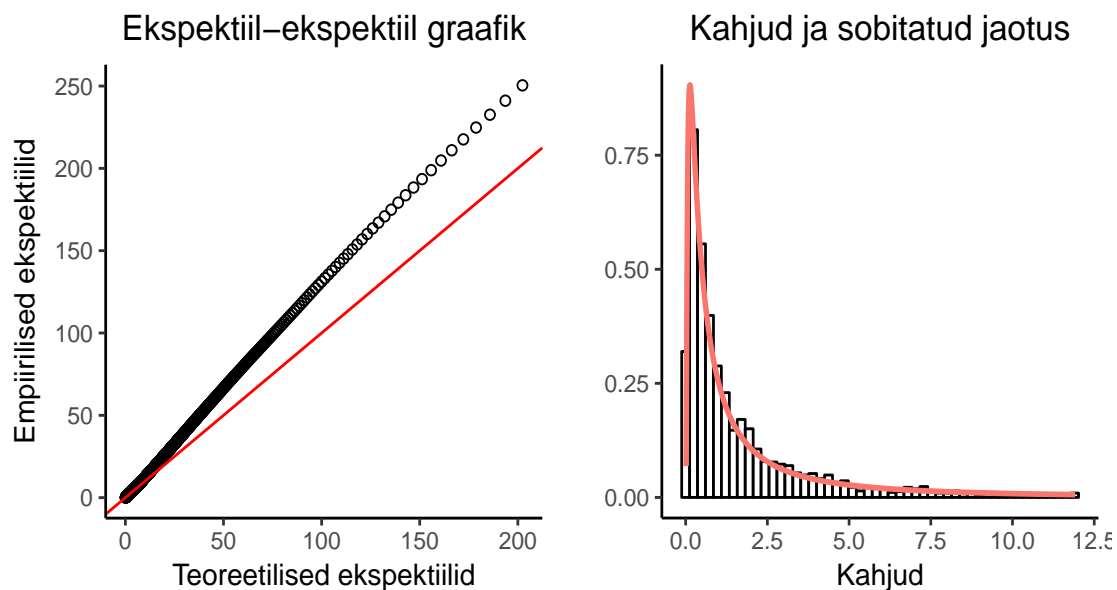
Joonis 10: Funktsiooni tulemus log-normaaljaotuse näitel

Näeme, et baasjuhul saame tulemuseks neli jaotuse sobivust kirjeldavat joonist, millest esimene on töö põhitulemus ehk valitud jaotusele, sellele hinnatud parameetritele ja andmestikule vastav ekspektiil-ekspektiil graafik.

Tutvustamaks funktsiooni ülejäänud parameetrite väärtustamist, toome näitena väljakutse, mille korral on kasutaja leidnud huvipakkuvana vaid ekspektiil-ekspektiil graafiku ning histogrammi. Paneme tähele, et ka ülejäänud parameetrid on määratud manuaalselt.

```
EEGraaf("log-gamma", andmed, hist = TRUE, kvant = FALSE,
        ekspKvant = FALSE, esimene = FALSE, cToke = 100,
        cParam = 1000, cSkaala = 1000, tArv = 50,
        aLimEksp = 0.995, aLimHist = 0.93)
```

Antud väljakutse tulemusena saame log-gamma jaotusele vastava ekspektiil-ekspektiil graafiku koos andmestikku kirjeldava tulpdiaagrammiga, millele on märgitud uuritava jaotuse tihedusfunktsioon.



Joonis 11: Funktsiooni tulemus log-gamma jaotuse näitel

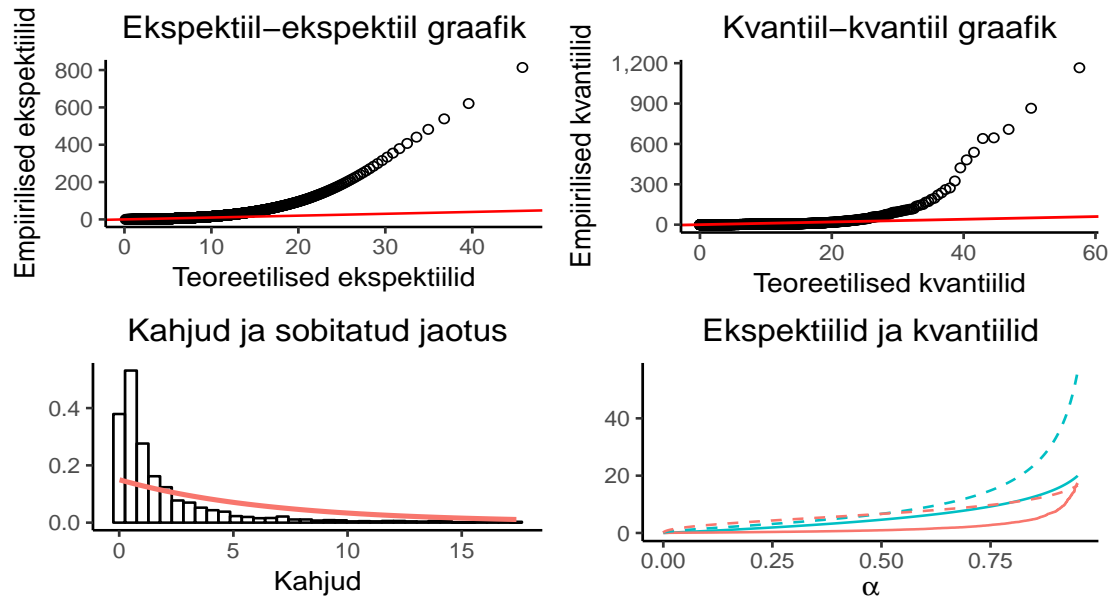
Graafikul on kaks jaotuse sobivust kirjeldavat joonist. Märkime, et histogrammil on eemaldatud jaotuse saba ning seetõttu ei paista parempoolselt joonisel välja empiirilise jaotuse erinevus teoreetilisest harvaesinevate väga suurte realisatsioonide korral.

## 2.2 Jaotuste sobivus testandmestikuga

Alapeatükis hindame vaadeldud jaotuste sobivust testandmestikuga, milles on 2797 kindlustuskahju väärtused. Minimaalne väärtus on 15,27, andmestiku suurimaks kahjuks on 1166171, keskmiseks kahju suuruseks on 6671,1 ja andmete mediaaniks on 952.97. Hindamaks, kas sobitatavad teoreetilised jaotused vastavad andmete empiirilisele jaotusele, kasutame kõiki loodud funktsiooni pakutavaid võimalusi. Põhiliselt keskendume jaotuse sobivuse hindamisel ekspektiil-ekspektiil graafikule ning selle võrdlemisele kvantiil-kvantiil graafikuga, kuid tulemusi visualiseerime ka histogrammi ja esimeses peatükis vaadeldud ekspektiilide ning kvantiilide muutumist kirjeldava joonisega. Kõik alapeatükis esitatavad graafikud on loodud eelnevalt kirjeldatud funktsiooni väljakutsetena. Telgedele märgitud kahjusuurused on vastavalt parameetri `cSkaala` vaikeväärtusele antud tuhandetes ühikutes.

## 2.2.1 Eksponentjaotus

Esmalt hindame eksponentjaotuse sobivust andmetega. Selleks alustame jaotuse sobivust kirjeldavast koondgraafikust. Eksponentjaotusele hinnatud parameetriks on  $\lambda = 1,5 \cdot 10^{-4}$ .

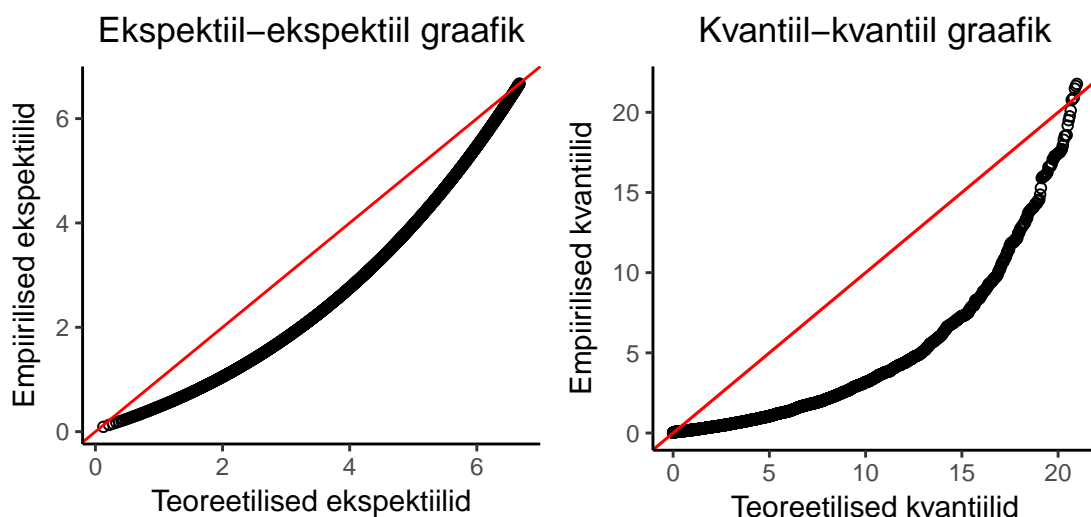


Joonis 12: Eksponentjaotuse sobivus andmetega

Ekspektiil-ekspektiil graafikult näeme, et suuremate kahjude korral paiknevad punktid ülalpool telgedevahelist sirget ehk teisisõnu empiirilised ekspektiilid ületavad teoreetilisi. See viitab teoreetilise jaotuse liiga kergele sabale ning annab alusta arvata, et antud jaotust kasutades alahindaksime andmete kirjeldamisel suurte kahjude esinemistõenäosust. Sama üldmulje jääb ka kvantiil-kvantiil graafikult, kuid võrreldes jooniseid omavahel, selguvad ekspektiil-ekspektiil graafikute erinevused võrreldes kvantiil-kvantiil joonistega. Esiteks on näha, et ekspektiilid hakkavad punasest sirgest kõrgemale tõusma kvantiilidest oluliselt varasemalt. See tuleneb ekspektiilide definitsioonist, mis arvestab iga argumendi  $\alpha$  väärtuse korral kõigi valimi elementidega proportsionaalselt nende väärtustega. Kvantiilid seevastu arvestavad ekstremaalsete realisatsioonidega vaid suhteliselt nende esinemiste osakaaluga ning seetõttu pööravad madalate  $\alpha$  väärtuste korral oluliselt vähem tähelepanu jaotuse sabale. Teine joonistevaheline erinevus on ekspektiil-ekspektiil graafiku punktide paigutuse siledus võrreldes kvantiilide punktiparvega. Kuna iga ekspektiili arvutamisel arvestatakse kõigi valimi elementide väärtustega, mõjutavad valimi omapärast tulenevad väiksemad kõikumised ekspektiilide graafikut kvantiilide joonisest vähemal määral. Kirjeldatud omadus võiks vähendada ohtu

teoreetilist jaotust andmestikule üle sobitada, kuid samas võib jätta piisava tähelepanuta ka põhjuslikud väiksemad omapärad andmete jaotuses. Histogrammilt on näha, et tihedusfunktsiooni graafik ei sobi tulpdiagrammiga üheski piirkonnas ning ka viimane graafik kinnitab nii teoreetiliste ja empiiriliste kvantiilide vahelist kui teoreetiliste ning valimiekspektilide vahelist erinevust.

Kuna ekspektil-ekspektil ja kvantiil-kvantiil graafikute puhul tundub, et punktiparv hakkab telgedevahelisest sirgest selgelt eemalduma alles suurte kahjude korral, vaatame eraldi ka väiksemate kahjude sobivust. Selleks piirame nii ekspektil-ekspektil kui kvantiil-kvantiil graafikule kantavaid statistikuid argumenti  $\alpha$  väärtuste kaudu. Teisisõnu eemaldame joonistelt jaotuste saba selliselt, et oleks näha, millise kahjude suuruse juures toimub teoreetilise ja empiirilise jaotuse eraldumine. Selleks kasutame funktsiooni parameetreid `aLimEks` ja `aLimKvant`. Piirangute seadmisel valime vähimad  $\alpha$  väärtused, mille korral on toimunud punktiparve pöördumatu eemaldumine telgedevahelisest sirgest. Täiendava joonise eesmärgiks on määrata, millise kahjusuuruse juures toimub andmete jaotuse eraldumine teoreetilisest jaotusest ning võrrelda ekspektil-ekspektil graafikute käitumist kvantiil-kvantiil joonistega.



Joonis 13: Eksponentjaotuse sobivus piiratud  $\alpha$  väärtuste korral

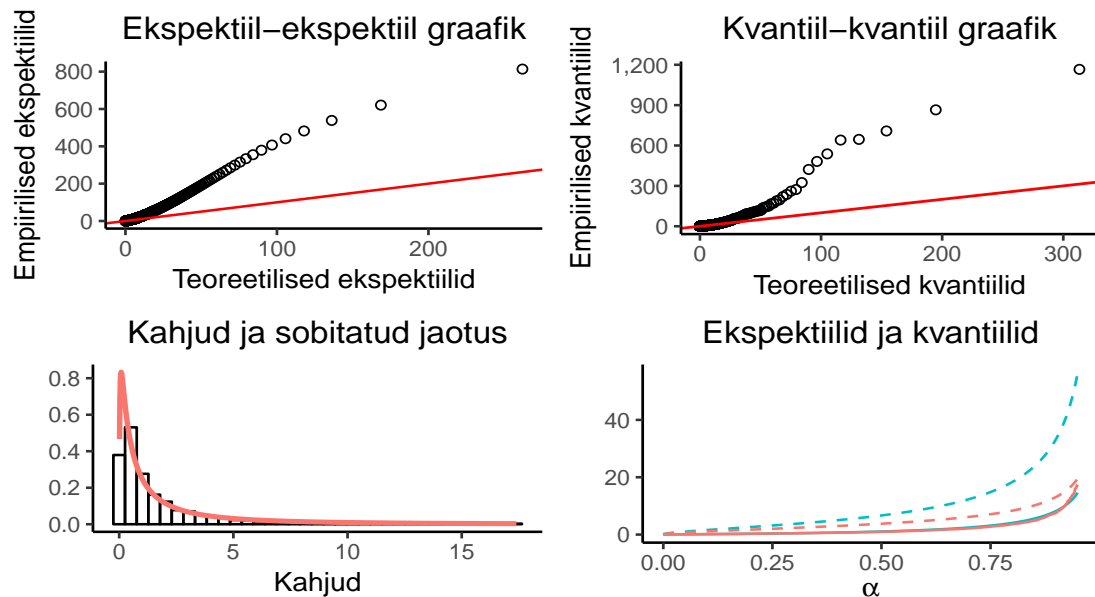
Joonisele kantud ekspektilidele vastavad  $\alpha$  väärtused on ülevalt piiratud arvuga 0,5 ehk teisisõnu on ekspektilide väärtuseid ülalt tõkestatud keskmise kahjusuurusega. Märkime taaskord, et keskmiseks kahjusuuruseks on 6671,1. Kvantiil-kvantiil graafikult on eemaldatud oluliselt väiksem osa kahjustest, piirates statistikutele vastavad argumenti  $\alpha$  väärtused ülalt arvuga 0,957. Näeme, et oodatult on empiirilised ja teoreetilised ekspektilid sarnased  $\alpha = 0,5$  korral, kuid ülejäänud osas erineb punktiparv telgedevahelisest



sirgest ka väiksemate kahjude korral. Kvantiile tähistavate punktide asetus meenutab kujult parabooli ning teoreetilised kvantiilid erinevad selgelt empiirilistest.

### 2.2.2 Log-normaaljaotus

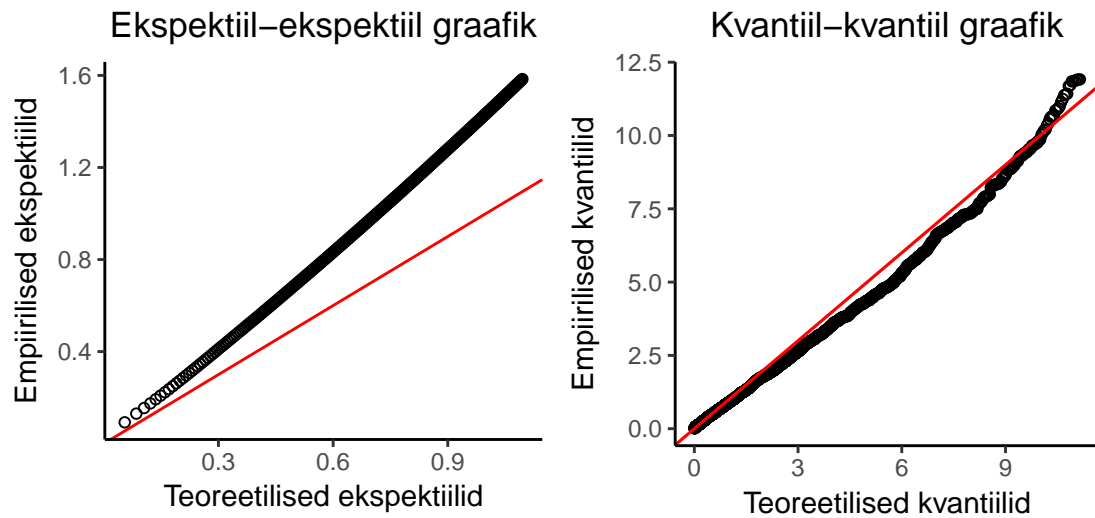
Uurime ekspektiilide abil jaotuse sobivust log-normaaljaotusega. Jaotusele vastavate parameetrite hinnanguks on  $\mu = 6,97$  ja  $\sigma^2 = 2,53$ .



Joonis 14: Log-normaaljaotuse sobivus andmetega

Näeme, et nii ekspektiil-ekspektiil kui kvantiil-kvantiil graafik paiknevad telgedevahele sirgest kõrgemal ehk teisisõnu nii empiirilised ekspektiilid kui empiirilised kvantiilid ületavad teoreetilisi statistikuid. See viitab, et ka log-normaaljaotuse saba on andmete jaotuse omast kergem ning ei arvesta piisavalt suurte kahjude esinemissagedusega. Lisaks täheldame taaskord ekspektiilide siledamat asetsemist graafikul. Histogrammilt näeme, et nii empiirilisel kui teoreetilisel jaotusel esineb enim väikeseid, kuid mitte nullilähedasi kahjusid ja samuti tunduvad jaotused sarnased ülejäänud osas. Viimasel graafikul paneme tähele ekspektiile tähistavate joonte erinevust suuremate  $\alpha$  väärtuste korral.

Piirame joonistele kantavate statistikute väärtuseid eesmärgiga määrata, millise kahju suuruse juures teoreetilised statistikud eralduvad empiirilistest.

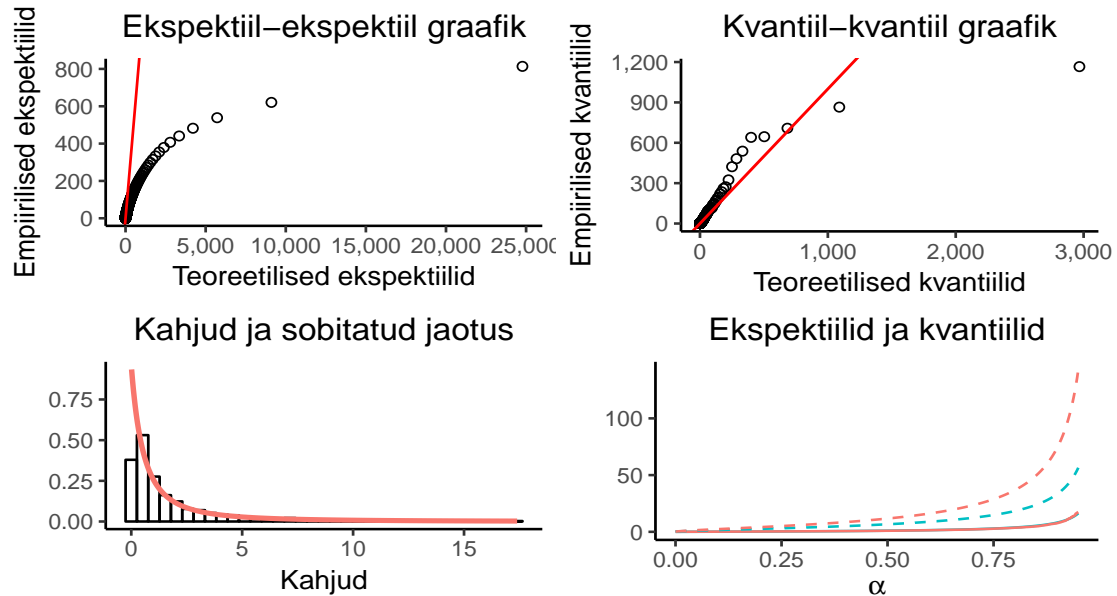


Joonis 15: Log-normaaljaotuse sobivus piiratud  $\alpha$  väärtuste korral

Kuna teoreetilised ekspektiilid eralduvad valimi põhjal arvutatutest koheselt, on ekspektiilide graafikul seatud  $\alpha \leq 0,1$ . Kvantiilid järgivad telgedevahelist sirget oluliselt suuremate väärtusteni ning nende puhul on valitud  $\alpha \leq 0,93$ . Joonised illustreerivad ilmekalt ekspektiil-ekspektiil ja kvantiil-kvantiil graafikute erinevat rõhuasetust jaotuse sobivuse hindamisel. Teisisõnu arvestab iga ekspektiil kogu valimit ning laseb end seetõttu mõjutada ka valimi ekstreemalsetest väärtustest, kuid kvantiil-kvantiil graafiku puhul võtame arvesse vaid kvantiilist suuremate väärtuste esinemissagedust. Kokkuvõtvalt näeme, et log-normaaljaotus sobib valdava enamikuga andmetest, kuid alahindab oluliselt suurkahjude esinemissagedust.

### 2.2.3 Pareto jaotus

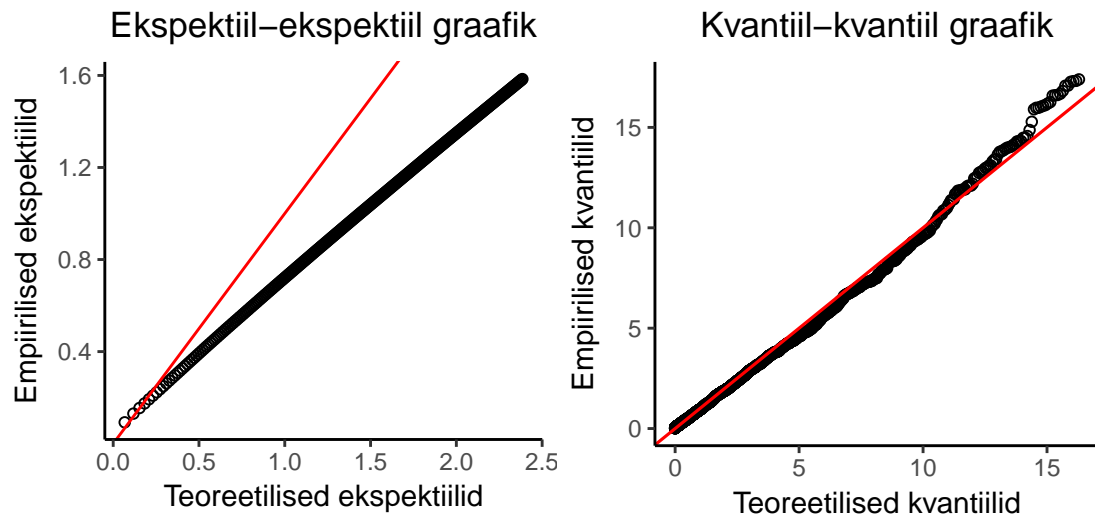
Hindame Pareto jaotuse sobivust andmetega. Pareto jaotuse parameetrite hinnanguteks on  $\gamma = 1,1$  ja  $\beta = 1141,9$ .



Joonis 16: Pareto jaotuse sobivus andmetega

Pareto jaotuse puhul paiknevad ekspektiil-ekspektiil graafiku punktid kogu joonise ulatuses allpool telgedevahelist sirget, kuid kvantiil-kvantiil graafiku puhul eralduvad teoreetilised statistikud empiirilistest alles väga suurte kahjude korral. Kuna teoreetilised ekspektiilid on suuremad neile vastavatest empiirilistest ekspektiilidest, võiks ekspektiil-ekspektiil graafiku põhjal arvata, et Pareto jaotus ülehindab kirjeldatavaid kahjusid. Kvantiil-kvantiil graafiku korral on jaotuste erinevus ebaselgem. Näeme, et suurel osal joonisest ületavad valimkvantiilid teoreetilisi kvantiile, kuid väga kõrgete argumenti  $\alpha$  väärtuste korral on hoopis teoreetilised kvantiilid suuremad. Tulpdiagrammilt näeme, et Pareto jaotus järgib valdavalt osa andmetest õnnestunult, kuid võib erineda empiirilise jaotusest väga väikeste kahjude korral. Nimelt on Pareto jaotuse tihedusfunktsioon kahanev kõigi mittenegatiivsete väärtuste korral, kuid andmete seas leidub kõige rohkem väikeseid, aga mitte nullilähedasi kahjusid. Viimasel joonisel märkame teoreetiliste ja valimiekspektiilide erinevust suurte  $\alpha$  väärtuste korral.

Vaatame ka Pareto jaotuse puhul, milliste väärtuste juures tekib teoreetiliste statistikute eraldumine empiirilistest.

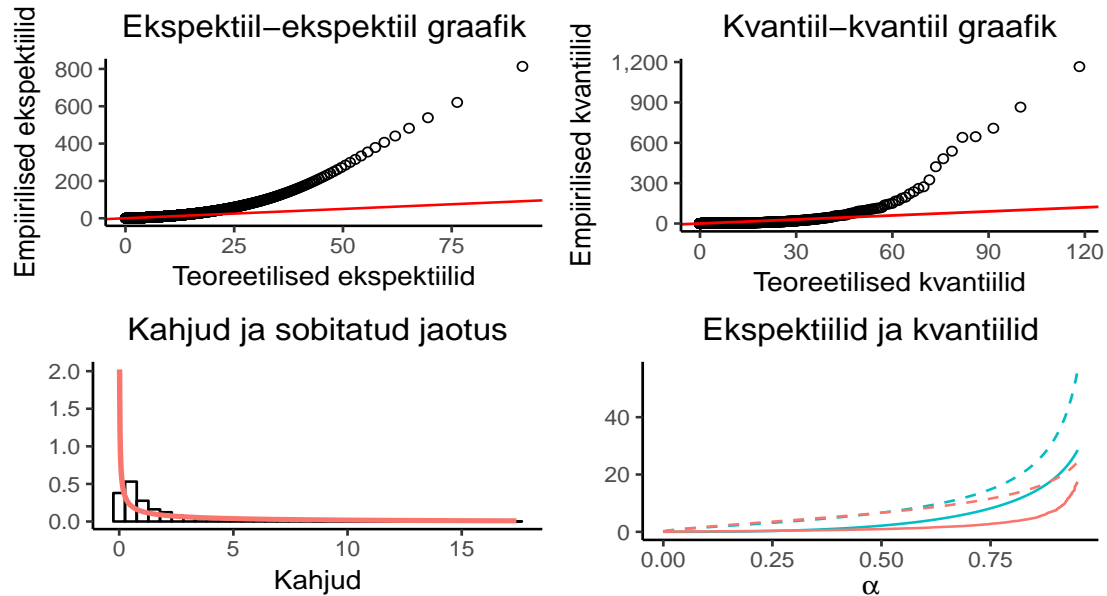


Joonis 17: Pareto jaotuse sobivus piiratud  $\alpha$  väärtuste korral

Joonisele kantud ekspektiilide jaoks on seatud  $\alpha \leq 0,1$  ja kvantiilidel  $\alpha \leq 0,95$ . Näeme, et ekspektiilid eralduvad telgedevahelisest sirgest taaskord koheselt, kuid kvantiilid järgivad punast joont suurel osal andmestikust. Nagu jooniselt 16 nägime, on Pareto jaotuse saba oluliselt raskem andmete jaotusest ning ekspektiilid arvestavad selle erinevusega ka väiksemate  $\alpha$  väärtuste korral määral, mis suurendab teoreetiliste statistike väärtuseid empiirilistega võrreldes. Teoreetilised kvantiilid jaotuse ekstremaalsete realisatsioonide väärtustega sellisel määral ei arvesta ning järgivad toodud osas valimi realisatsiooni õnnestunult. Kokkuvõtvalt sobitub ka Pareto jaotus valdava enamikuga kahjust, kuid ülehindab selgelt suurkahjude osakaalu.

## 2.2.4 Gammajaotus

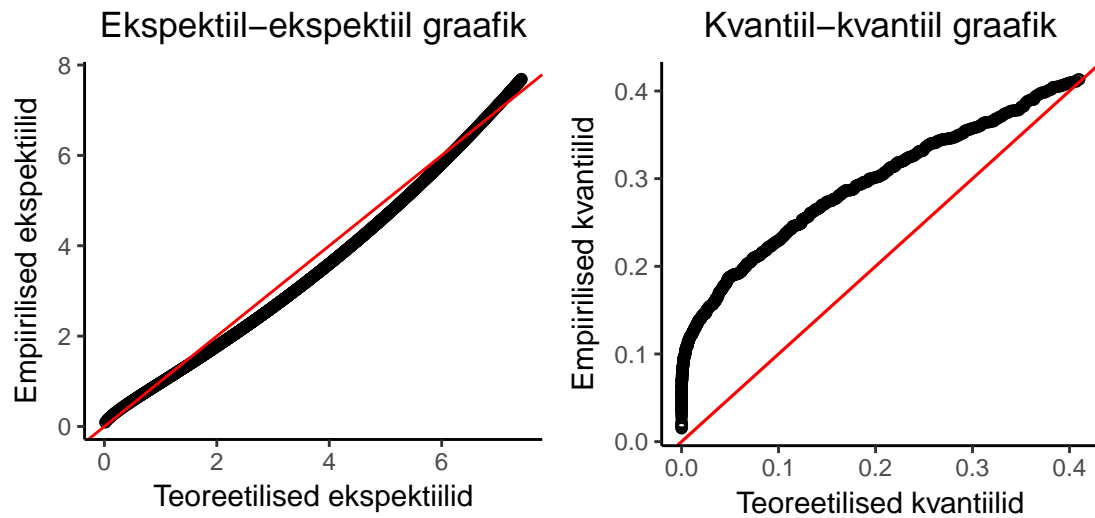
Uurime graafikute abil kahjude sobivust gammajaotusega. Jaotusele hinnatud parameetriteks on  $\gamma = 0,36$  ja  $\beta = 5,47 \cdot 10^{-5}$ .



Joonis 18: Gammajaotuse sobivus andmetega

Esimeselt kahelt graafikult näeme, et nii ekspektiil-ekspektiil kui kvantiil-kvantiil graafiku puhul on punktiparv suuremate  $\alpha$  väärtuste korral telgedevahelisest sirgest kõrgemal, kuid ekspektiilid eristuvad sirgest kvantiilidega võrreldes väiksemate  $\alpha$  väärtuste korral. Histogrammi põhjal tundub, et gammajaotuse tihedusfunktsioon ületab tulpade kõrguseid väga väikeste kahjude korral ja jääb tulpadest madalamaks väikeste, kuid mitte nullilähedaste kahjude andmestiku elementide korral. Viimaselt jooniselt näeme, et teoreetilised statistikud eristuvad empiirilistest selgemalt suuremate  $\alpha$  väärtuste korral.

Piirame joonistele kantavate statistikute väärtuseid eesmärgiga määrata, millise kahju suuruse juures gammajaotusele vastavad teoreetilised statistikud eralduvad empiirilistest.

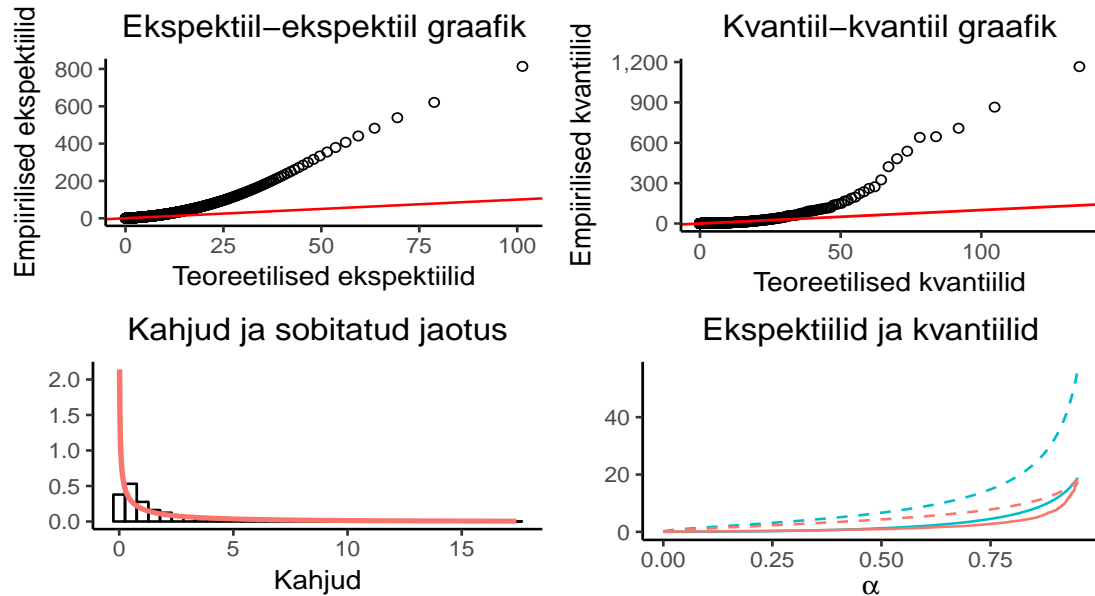


Joonis 19: Gammajaotuse sobivus piiratud  $\alpha$  väärtuste korral

Ekspektiil-ekspektiil graafikul on seatud  $\alpha \leq 0,55$  ja kvantiil-kvantiil joonisel on  $\alpha \leq 0,28$ . Näeme, et teoreetilised ekspektiilid sarnanevad ligi poolel andmestikust empiirilistega. Kuna hinnatud parameetrite korral on gammajaotuse keskväärus sarnane keskmise kahjusuurusega, on ekspektiilide lõikekoht ligikaudu argumenti väärtuse  $\alpha = 0,5$  juures. Kvantiil-kvantiil graafikul paistab silma punktisarve selge eristumine telgedevahelisest sirgest.

### 2.2.5 Weibulli jaotus

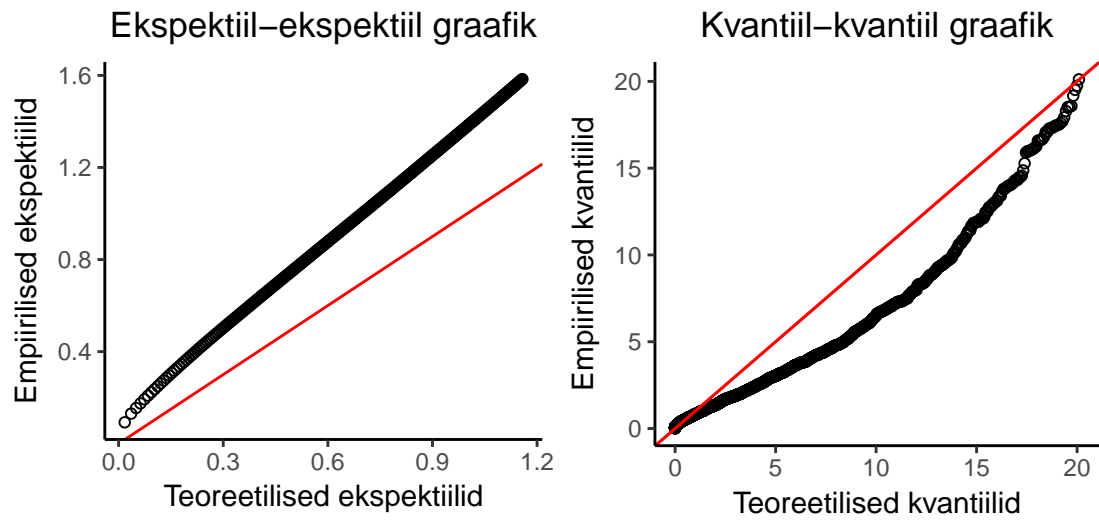
Vaatame andmete sobivust Weibulli jaotusega. Kasutatud on hinnatud parameetreid  $k = 0,54$  ja  $\lambda = 2462,2$ .



Joonis 20: Weibulli jaotuse sobivus andmetega

Weibulli jaotusele vastavad graafikud sarnanevad gammajaotuse joonistega. Esimesel kahel graafikul ületavad empiirilised statistikud suurte  $\alpha$  väärtuste korral teoreetilisi. Teisisõnu on Weibulli jaotus sarnaselt gammajaotusega andmete jaotusest kergema sabaga ning ei arvesta seetõttu piisavalt suurt kahjudega. Paneme tähele, et selge eristumine telgedevahelisest sirgest tekib esmapilgul ekspektiilide korral kvantiilidest varasemalt, kuid kontrollime tulemust hilisemalt jaotuse algosale keskendunud joonisega. Tulpdiagrammilt näeme teoreetilise ja valimijaotuse vahelist erinevust ka väikeste kahjude korral ning viimaselt jooniselt teoreetiliste ja empiiriliste ekspektiilide erinevaid väärtuseid suurte argumenti  $\alpha$  väärtuste korral.

Piirame Weibulli jaotusele vastavatele joonistele kantavate statistikute argumenti  $\alpha$  väärtuseid.



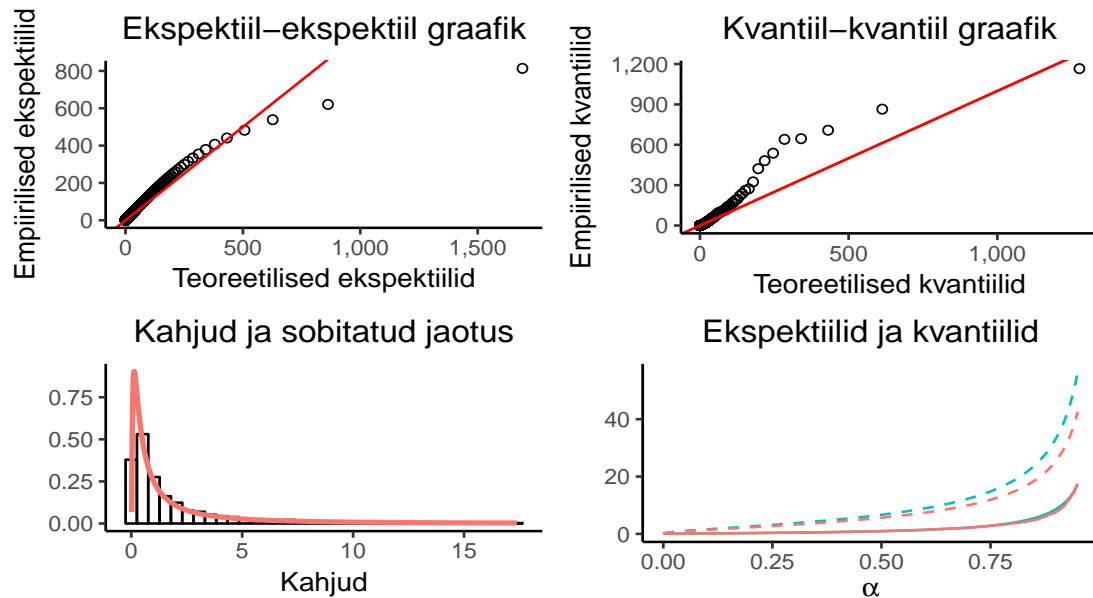
Joonis 21: Weibulli jaotuse sobivus piiratud  $\alpha$  väärtuste korral

Ekspektiilide joonisel on  $\alpha$  väärtuseid piiratud arvuga 0,1 ja kvantiilide graafikul arvuga 0,955. Nii ekspektiil-ekspektiil kui kvantiil-kvantiil graafikult paistab empiiriliste ja teoreetiliste statistikute omavaheline erinevus ning sarnaselt gammajaotusega peame jääma otsusele, et Weibulli jaotus ei sobi andmetega.



### 2.2.6 Log-gamma jaotus

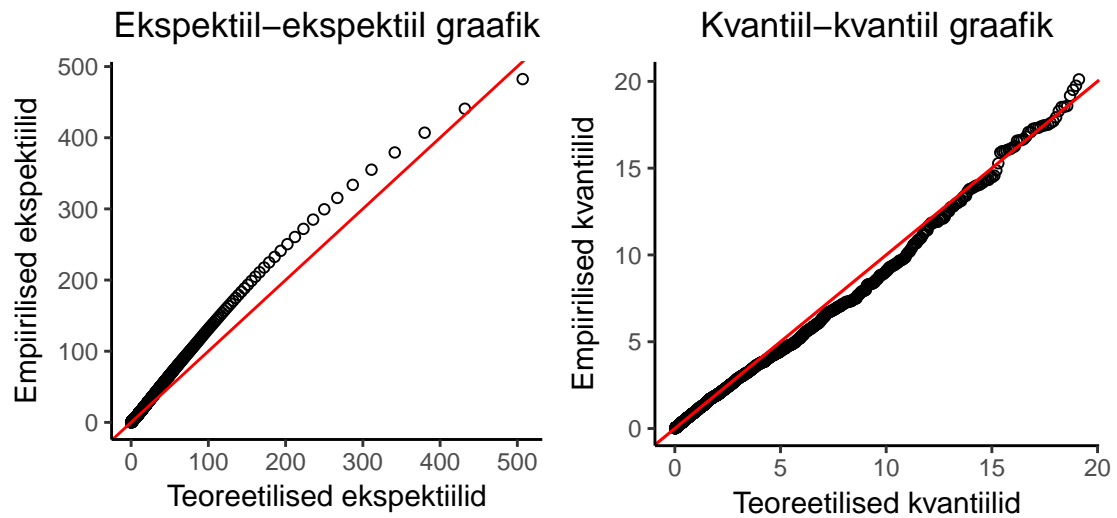
Viimaks uurime, kas andmed sobivad log-gamma jaotusega. Parameetrite hinnanguks on  $\gamma = 2,79$  ja  $\beta = 19,44$ .



Joonis 22: Log-gamma jaotuse sobivus andmetega

Esimeselt kahelt jooniselt näeme, et log-gamma jaotuse puhul on teoreetilised statistikud empiirilistele näitajatele teistest jaotustest sarnasemad. Ekspektiil-ekspektiil graafik järgib telgedevahelist sirget pea kogu andmestikul ja erinevus andmetega kerkib seejuures esile vaid kolme suurima kahju korral. Kvantiil-kvantiil graafikul on teoreetilise ja empiirilise jaotuse vaheline erinevus märgatav suuremas osas, kuid seejuures asub telgedevahelisel sirgel ka andmestiku suurim kahju. Histogrammilt näeme, et log-gamma jaotuse tihedusfunktsiooni ja tulpade kõrguste vahel on ainus silmapaistev erinevus tihedusfunktsiooni maksimumkoha läheduses, kuid see võib tuleneda ka histogrammi tulpade piiratud arvust ning ei pruugi seetõttu veel viidata jaotuste erinevusele. Muus osas tundub tihedusfunktsioon järgivat histogrammi kuju. Jaotuse sobivust kinnitab ka viimane graafik. Nimelt on ekspektiilide ja kvantiilide graafikult märgatav mõningane erinevus küll teoreetiliste ja empiiriliste ekspektiilide vahel suuremate  $\alpha$  väärtuste korral, kuid kvantiile tähistavad jooned on vaadeldavas osas äärmiselt sarnased.

Jätame graafikult välja telgedevahelisest sirgest selgelt eralduvad punktid. Eesmärgiks on jõuda selguseni, millise kahjusuuruse juures toimub teoreetilise ja empiirilise jaotuse eraldumine ning anda täpsem joonis jaotuse algusosa sobivusest.



Joonis 23: Log-gamma jaotuse sobivus piiratud  $\alpha$  väärtuste korral

Ekspektiilide graafikul on seatud  $\alpha \leq 0,999$ . Näeme, et ekspektiilid on telgedevahelise sirge vahetus läheduses valdava enamiku kahjusuuruste korral ning vaid 0,1% andmestiku suurimatele kahjudele vastavatest empiirilistest ekspektiilidest erinevad selgelt teoreetilistest ekspektiilidest. Kvantiilide jooniselt näeme, et teoreetiline jaotus sobib empiirilisega kuni umbes kahjusuuruseni 20000. Kokkuvõtvalt järgib log-gamma jaotus andmestikku õnnestunult valdava enamiku kahjude korral ning erinevus tekib vaid mõnele suurkahjudele vastavate statistikute vahel. Kindlust jaotuse parima sobivuse osas lisab asjaolu, et vastavalt joonisele [22](#) sobivad teoreetilised ja empiirilised kvantiilid ka andmestiku suurima kahju korral.

## Kokkuvõte

Käesolevas Bakalaureusetöös tutvustati ekspektiile ning uuriti nende abil jaotuse sobivuse hindamist ekspektiil-ekspektiil graafikute näol.

Analoogia tõttu anti esimeses peatükis ülevaate kvantiilidest ja kvantiil-kvantiil graafikutest ning toodi sisse üldistatud kvantiili mõiste. Üldistatud kvantiilide erijuhuks on lisaks laialt levinud kvantiilidele ka töös kesksel kohal olevad statistikud - ekspektiilid. Töös toodi ekspektiilide definitsioon, käsitleti ekspektiilide omadust, mille kohaselt juhusliku suuruse  $X$  0,5-ekspektiil on võrdne juhusliku suuruse keskväärtusega,  $e_{0,5}(X) = EX$  ning tõestati see. Seejärel anti ekspektiilide leidmise võrrandid levinud kahjujaotuste jaoks ning avaldati ekspektiilide leidmise võrrand log-gamma jaotuse jaoks. Kirjeldamaks vaadeldud jaotustele vastavate ekspektiilide muutumist jaotuste erinevate parameetrite korral ning võrdlemaks ekspektiile kvantiilidega, visualiseeriti statistikuid joonistel. Esimest peatükki lõpetab näide ekspektiilide leidmisest diskreetse juhusliku suuruse korral.

Teises peatükis loodi esmalt ekspektiil-ekspektiil graafikute realiseerimiseks vajalik tarkvara ning selgitati selle tööpõhimõtet ja kasutamist. Seejärel rakendati loodud funktsiooni reaalsele kahjuandmetele, hindamaks kõigi käsitletud kahjujaotuste sobivust andmetega. Analüüsi tulemusel selgus, et andmestikuga sobitub vaadeldud jaotustest kõige paremini log-gamma jaotus.

Bakalaureusetöö eesmärgiks oli tutvustada ekspektiile ning realiseerida graafiline meetod ekspektiilide abil jaotuse sobivuse hindamiseks. Autor hindab eesmärgi täidetuks ning usub, et kirjeldatud metoodika võimaldab ebasümmeetriliste jaotuste korral pöörata kvantiil-kvantiil graafikutest rohkem tähelepanu jaotuste ekstremaalsetele realiseerimistele. Lisaks võimaldab ekspektiil-ekspektiil graafikute kasutuselevõtt tulenevalt punktisarve siledamast paigutusest vältida teoreetilise jaotuse andmetele ülesobitamist.

Tööle lisab praktilist väärtust loodud rakendus jaotuse sobivuse hindamiseks, mis võimaldab kasutajal otsuse langetamiseks kasutada nelja tüüpi jooniseid:

- ekspektiil-ekspektiil graafik;
- kvantiil-kvantiil graafik;
- normeeritud histogramm koos sobitatava jaotuse tihedusfunktsiooniga;
- teoreetiliste ja empiiriliste ekspektiilide ning kvantiilide muutumist kirjeldav graafik.

## Viited

- [1] R. P. Brent, *Algorithms for Minimization without Derivatives*. Prentice-Hall, Englewood Cliffs N.J, 1973.
- [2] I. S. Gradshteyn, I. M. Ryzhik, *Table of Integrals, Series and Products, Seventh Edition*. Elsevier, San Diego, 2007.
- [3] A. Gusnanto *Statistical Modelling*. Loengukonspekt, Leeds, 2011. <http://www1.maths.leeds.ac.uk/arief/MATH2735/qplots.html> [15.04.2017].
- [4] T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning, 2nd edition*. Springer, California, 2009.
- [5] R. V. Hogg, S. A. Klugman, *Loss Distributions*. John Wiley & Sons, Iowa, 1984.
- [6] A. Kaasik, M. Käärik, *Distributions in Financial Mathematics*. Loengukonspekt, Tartu, 2015.
- [7] S. A. Klugman, H. H. Panjer, G. E. Willmot, *Loss Models From Data to Decisions*. John Wiley & Sons, New Jersey, 2004.
- [8] R. Koenker, *Quantile Regression*. Cambridge University Press, New York, 2005.
- [9] T. Leiger, *Matemaatiline analüüs III*. Loengukonspekt, Tartu, 2016.
- [10] H. Pishro-Nik, *Introduction to Probability, Statistics and Random Processes*. Kappa Research, USA, 2014.
- [11] H. Puksand, *Ekspektiilid ja nende kasutamine riskimõõduna*. Magistritöö, Tartu, 2015.
- [12] K. Pärna, *Tõenäosusteooria algkursus*. Tartu Ülikooli Kirjastus, Tartu, 2013.
- [13] L. Rosasco, E. De Vito, A. Caponnetto, M. Piana, A. Verri, *Are Loss Functions All the Same?* Italia, 2003.
- [14] I. Tammeraid, *Matemaatiline analüüs I*. TTÜ Kirjastus, Tallinn, 2001.

## Lisad

### Funktsiooni **EEGraaf** lähtekood

```
library(zipfR)#mittetaielik gammafunktsioon
library(expint)#mittetaielik gamma
library(actuar)#log-gamma jaotus
library(fitdistrplus)#parameetrid
library(ggplot2)
library(gridExtra)#jooniste koos esitamine
require(scales)#joonistel teadusliku notatsiooni valtimiseks

#Funktsioon, mis:

#1) Leiab teoreetilised ekspektilid valitud hüpoteesjaotusele
#vastavalt. Selleks:
#defineerime kõik toos kasitletud vorrandid;
#hindame kasutaja valitud jaotusele vastavad parameetrid ST
#meetodil fitdistributionplus paketi abil;
#leiame andmestiku pikkusele vastava arvu ekspektille selleks
#vajalikku vorrandit iteratiivselt (igale alphale vastavalt,
#seega andmestiku ridade arv kordi) igal iteratsioonisammul
#numbriliselt (Brenti meetod) lahendades (funktsiooni absoluut-
#vaartuse minimeerimine).

#2) Leiab andmestikule vastavad empiirilised ekspektilid
#definiitsioonil pohinevalt. Selleks itereerime üle koigi alphade
#(koigi valimi elementide), kus minimeerime funktsiooni,
#mis itereerib omakorda üle koigi elementide. Minimiseerimiseks
#kasutame taaskord Brenti meetodit.

#3) Joonistab hajuvusdiagrammi, mille x-teljel on teoreetilised
#ning y-telje empiirilised ekspektilid. Teisisõnu on funktsiooni
#too tulemuseks ekspektil-ekspektil graafikud.

EEGraaf = function(jaotus, andmestik, cToke = 1e3, cParam = 1e4,
                    hist = TRUE, kvant = TRUE, ekspKvant = TRUE,
                    tArv = 35, esimene = TRUE, cSkaala = 1000,
                    aLimHist = 0.95, aLimKvant = 1, aLimEksp = 1){

  #PARAMEETRID:
```

```

#jaotus - kasutaja hupoteesjaotus. Peab sisalduma vektoris
#c("eksponent", "log-norm", "Pareto", "gamma", "Weibull",
# "log-gamma")

#andmestik - uuritav andmestik vektorina (elemendid klassist
#"numeric")

#cToke - arv, mida kasutatakse ekspektiili ulalt tokestava
#kvantiili skaleerimisel. Toke = kvantiil*c + c

#cParam - arv, millega viiakse labi skaalateisendus. Teisisonu
#jagatakse andmestik elementviisiliselt cParam-ga ja
#manipuleeritakse muudatusele vastavalt hinnatud parameetreid,
#et need sobiksid algsele andmestikule.

#hist - TRUE/FALSE. kas joonistame normeeritud histogrammi
#andmestikust, millele on margitud hinnatud parameetritega
#hupoteesjaotuse tihedusfunktsioon

#kvant - TRUE/FALSE. kas joonistame hinnatud parameetritega
#hupoteesjaotusele vastava kvantiil-kvantiil graafiku

#ekspKvant - TRUE/FALSE. kas joonistame empiiriliste ning
#teoreetiliste ekspektiilide-kvantiilide graafiku

#tArv - naturaalarv, Seab histogrammi tulpade arvu

#aLimEksp - arv. maarab, kui suurt osa andmestiku algusest
#visualiseerime ekspektiil-ekspektiil graafikul

#aLimKvant - arv. maarab, kui suurt osa andmestiku algusest
#visualiseerime kvantiil-kvantiil graafikul

#aLimHist - arv. maarab, kui suurt osa andmestiku algusest
#kasutame histogrammil ja ekspektiilide-kvantiilide graafikul

#esimene - TRUE/FALSE. kas arvutame empiirilised ekspektiilid
#andmestikust ja salvestame tulemuse voi laeme eelnevalt
#salvestatud ekspektiilid

```

```

#cSkaala - arv, mis lubab teisendada skaalat loodavatel
#graafikutel, vaikevaartus 1000.

jaotused = c("eksponent", "log-norm", "Pareto", "gamma",
             "Weibull", "log-gamma")

#kontrollime, kas kasutaja sisestas lubatava jaotuse
if(jaotus %in% jaotused == FALSE){
  print("Sisestasid tundmatu jaotuse")
  return(NULL)
}

#DEFINEERIME VÕRRANDID
f_eksponent = function(par, alpha, e){#vorrant eksponentjaotuse
  #jaoks (absoluutvaartus lahendamiseks minimeerimise kaudu)
  abs((2*alpha-1)*exp(-par*e)/par+(alpha-1)*(e-1/par))
  #absoluutvaartus, et leida minimeerides nullkoht
}

f_lnorm = function(mu, sigma, alpha, e){#ekspektiilide
  #leidmise vorrand
  abs((1-2*alpha)*exp(mu+(sigma^2)/2)
      *pnorm((log(e)-sigma^2-mu)/sigma)
      + (2*alpha-1)*e*pnorm((log(e)-mu)/sigma)
      + alpha*(exp((mu+(sigma^2)/2))-e)) #absoluutvaartus,
  #et leida minimeerides nullkoht
}

f_pareto = function(e, g, b, a){#ekspektiilide vorrand
  abs((2*a-1)*((b+e)^(-g)*(g*e+b)*b^g)/(g-1)
      + (1-2*a)*e*b^g*(b+e)^(-g)+(a-1)*(e-b/(g-1)))
  #absoluutvaartus, et leida minimeerides nullkoht
}

f_gamma = function(e, g, b, alpha){#ekspektiilide vorrand
  abs((1-2*alpha)*e*gamma(a = g, x = b*e)/gamma(g)
      + (2*alpha - 1)*gamma(a = g+1, x = b*e)/(b*gamma(g))
      + (alpha-1)*(e-g/b)) #absoluutvaartus,
  #et leida minimeerides nullkoht
}

```

```

f_weibull = function(k, l, a, e){#ekspektilide vorrand,
  #l - lambda, a - alpha, e - ekspektil
  abs((2*a-1)*l*Igamma(a=(1+1/k), x = (e/l)^k, lower=FALSE)
    + (1-2*a)*e*exp(-(e/l)^k)+(a-1)*(e-l*gamma(1+1/k)))
  #absoluutvaartus, et leida minimeerides nullkoht
}

f_lgamma = function(e, g, b, alpha){#ekspektilide vorrand,
  #g - gamma, b - beeta, a - alpha, e - ekspektil
  abs((alpha*g^b/gamma(b)
    *(gammainc(a = b, x = (g-1)*log(e))/(g-1)^b
      - e*gammainc(a = b, x = g*log(e))/g^b))
    + (alpha-1)*g^b/gamma(b)
    *(e*Igamma(a = b, x = g*log(e), lower = TRUE)/g^b
      - Igamma(a = b, x = (g-1)*log(e), lower = TRUE)
        /(g-1)^b)) #absoluutvaartus, et leida
  #minimeerides nullkoht
}

vorrant = function(e, alpha, par, jaotus){
  if (jaotus == "eksponent"){#tingimus valib hilisemaks
    #kasutamiseks tihedusfunktsiooni
    f_eksponent(par = par, alpha = alpha, e = e)
  }else if (jaotus == "log-norm"){
    f_lnorm(mu = par[1], sigma = par[2], alpha = alpha, e = e)
  }else if (jaotus == "Pareto"){#vajab actuar paketti
    #ning parameetrite algvaartustamist
    f_pareto(e = e, g = par[1], b = par[2], a = alpha)
  }else if (jaotus == "gamma"){
    f_gamma(e = e, g = par[1], b = par[2], alpha = alpha)
  }else if (jaotus == "Weibull"){
    f_weibull(k = par[1], l = par[2], a = alpha, e = e)
  }else if (jaotus == "log-gamma"){#vajab actuar paketti
    #ning parameetrite algvaartustamist
    f_lgamma(e = e, g = par[1], b = par[2], alpha = alpha)
  }
}

n = length(andmestik)

```



```

#esialgu kasutame samasid alphade vaartuseid,
#mis kvantiilide puhul
e_teor = data.frame(((1:n)-0.5)/n, rep(NA, n))

colnames(e_teor) = c("alpha", "teor")

#HINDAME HUPOTEESJAOTUSE PARAMEETRID, neile vastavad
#ekspektiilide kvantiilidel pohinevad ulemised ja
#jaotusele vastavad alumised tokked

if (jaotus == "eksponent"){#tingimus valib hilisemaks
  #kasutamiseks tihedusfunktsiooni
  param=fitdist(andmestik/cParam, "exp")
  #eksponentjaotuse korral tuleb jagada.
  #sama probleem Helise toos
  alumine = 0
  par = c(param$estimate[1]/cParam)
  ulemine = qexp(p = e_teor[, 1], rate = par[1])*cToke + cToke
}else if (jaotus == "log-norm"){
  param=fitdist(andmestik, "lnorm")
  alumine = 0
  par = c(param$estimate[1], param$estimate[2])
  ulemine = qlnorm(p = e_teor[, 1], meanlog = par[1],
                    sdlog = par[2])*cToke + cToke
}else if (jaotus == "Pareto"){#vajab actuar paketti
  #ning parameetrite algvaartustamist
  param=fitdist(andmestik, "pareto",
                start = list(shape = 1, scale = 1))
  alumine = 0
  par = c(param$estimate[1], param$estimate[2])
  ulemine = qpareto(p = e_teor[, 1], shape = par[1],
                    scale = par[2])*cToke + cToke
}else if (jaotus == "gamma"){
  param=fitdist(andmestik/cParam, "gamma")
  alumine = 0
  par = c(param$estimate[1], param$estimate[2]/cParam)
  ulemine = qgamma(p = e_teor[, 1], shape = par[1],
                   rate = par[2])*cToke + cToke
}else if (jaotus == "Weibull"){

```

```

param=fitdist(andmestik, "weibull")
alumine = 0
par = c(param$estimate[1], param$estimate[2])
ulemine = qweibull(p = e_teor[, 1], shape = par[1],
                    scale = par[2])*cToke + cToke
}else if (jaotus == "log-gamma"){#vajab actuar paketti
  #ning parameetrite algvaartustamist
  param=fitdist(andmestik, dlgamma, start = list(ratelog = 1,
                                                  shapelog = 1))

  alumine = 1
  par = c(param$estimate[1], param$estimate[2])
  ulemine = qlgamma(p = e_teor[, 1], ratelog = par[1],
                    shapelog = par[2])*cToke + cToke
}

if (length(par) == 1){
  print(paste("Jaotusele", jaotus,
              "ja sisestatud andmestikule vastavaks",
              " parameetri hinnanguks on", par[1]))
}else{
  print(paste("Jaotusele", jaotus, "ja sisestatud andmestikule",
              "vastavaks parameetrite hinnanguks on", par[1],
              " ja ", par[2]))
}
#LEIAN TEOREETILISED EKSPEKTIILID
#Vektortehet ei saa teha, sest optim ei toeta seda.

for(i in seq(1, length(e_teor[,1]))) {
  opt = optimize(vorrand, lower = alumine, upper = ulemine[i],
                 par = par, alpha = e_teor[i, 1],
                 jaotus = jaotus)
  e_teor[i, 2] = opt$minimum
}

#EMPIIRILISED EKSPEKTIILID

#funktsioon, mis itereerib üle valimi ning leiab halbe antud
#ekspektiili vaartuse korral (definitsoonis minimeeritava
#vorrandi vaartus ilma keskmistamiseta)
efun = function(x, alpha, valim){

```

```

halve = 0
for(j in 1:length(valim)){
  halve = halve + alpha*(max(valim[j]-x,0))^2
  +(1-alpha)*(max(x-valim[j],0))^2
}
return(halve)
}

edef = function(a = min(valim), b = max(valim), valim){
  # x = ekspektiili vaartus (minimeerimisülesande
  #lahendamisel muudetav argument),
  #a = itereerimispiirkonna alguspunkt
  #b = itereerimispiirkonna lõpp-punkt
  df = data.frame()

  alpha = ((1:length(valim))-0.5)/length(valim)
  #ühtlaselt jaotatud ekspektiilid. analoog kvantiilidega.
  #Kas vajalik või võib harvendada?
  df_eksp = data.frame(alpha, emp = rep(NA, length(alpha)))
  #itereerime üle koigi alphade (koigile valimi elementidele
  #vastavate alphade), kus minimeerime funktsiooni,
  #mis itereerib üle koigi elementide.
  #selleks kasutame taaskord Brenti meetodit
  for(i in 1:length(alpha)){
    opt = optimize(efun, lower = a, upper = b,
                   alpha = alpha[i], valim = valim)
    df_eksp[i,2] = opt$minimum
  }
  df = df_eksp
  return(df)
}

#juhul, kui arvutame ekspektiile esimest korda,
#salvestame tulemuse, teisel juhul laeme varasemalt
#salvestatu

if (esimene == TRUE){
  e_eksp = edef(valim = andmestik)#andmestik, milles igale
  #ühtlaselt jaotatud alphale vastav ekspektiil
  saveRDS(e_eksp, file="andmestiku_EmpEksp.rds")
}

```

```

}else{
  e_eksp = readRDS("andmestiku_EmpEksp.rds")
}

ekspektiilid = merge(e_teor, e_eksp, by = c("alpha"))
colnames(ekspektiilid) = c("alpha", "teor", "emp")

#EKSPEKTIIL-EKSPEKTIIL GRAAFIK

#teeme hajuvusdiagrammi
gEksp = ggplot(ekspektiilid[ekspektiilid$alpha <= aLimEksp,],
               aes(x = teor/cSkaala, y = emp/cSkaala))
#loome tausta
gEksp = gEksp + theme_classic()
gEksp = gEksp + geom_point(shape=1) # ringikesed teeme
#seest tühjad
gEksp = gEksp + ylab("Empiirilised ekspektiilid")
gEksp = gEksp + xlab("Teoreetilised ekspektiilid")
gEksp = gEksp + ggtitle("Ekspektiil-ekspektiil graafik")
gEksp = gEksp + theme(plot.title = element_text(hjust = 0.5))
#pealkiri keskele
gEksp = gEksp + geom_abline(slope = 1, color = "red",
                             linetype = "solid", size = 0.5)
gEksp = gEksp + scale_y_continuous(labels = comma)
gEksp = gEksp + scale_x_continuous(labels = comma)

#HISTOGRAMM (kasutaja soovi korral, koos hupoteetilise
#jaotuse tihedusega) ja
#KVANTIIL-KVANTIIL GRAAFIK (kasutaja soovil)
#EKSPEKTIILIDE JA KVANTIILIDE GRAAFIK (kasutaja soovil,
#teoreetilised ja empiirilised ekspektiilid ning kvantiilid)
if((hist == TRUE) || (kvant == TRUE) || (ekspKvant == TRUE)){

  #kuna tegemist vaga pika sabaga jaotusega, jatame suurimad
  #kahjud histogrammilt ara
  andmestik = data.frame(andmestik)
  sort = andmestik[with(andmestik, order(andmestik)), ]

  algus = head(sort, n = aLimHist*length(sort))

```

```

#Lisamaks histogrammile vastava jaotuse tihedusfunktsiooni
#graafikut, leiame selle vaartused:
x_arg = seq(min(algus), max(algus), length = 1000)
#argumendid

#jargnevas arvutame ka kasutatud alphadele ja hinnatud
#jaotusele vastavad kvantiilid (koodi lihtsuse huvidus
#ning madalast (ajalisest) keerukusest tulenevalt)
if (jaotus == "eksponent"){#tingimus valib hilisemaks
  #kasutamiseks tihedusfunktsiooni
  fx = dexp(x_arg, rate = par[1])
  qx = qexp(e_teor[,1], rate = par[1])
}else if (jaotus == "log-norm"){
  fx = dlnorm(x_arg, meanlog = par[1], sdlog = par[2])
  qx = qlnorm(e_teor[,1], meanlog = par[1], sdlog = par[2])
}else if (jaotus == "Pareto"){#vajab actuar paketti
  #ning parameetrite algvaartustamist
  fx= dpareto(x_arg, shape = par[1], scale = par[2])
  qx = qpareto(e_teor[,1], shape = par[1], scale = par[2])
}else if (jaotus == "gamma"){
  fx = dgamma(x_arg, shape = par[1], rate = par[2])
  qx = qgamma(e_teor[,1], shape = par[1], rate = par[2])
}else if (jaotus == "Weibull"){
  fx = dweibull(x_arg, shape = par[1], scale = par[2])
  qx = qweibull(e_teor[,1], shape = par[1], scale = par[2])
}else if (jaotus == "log-gamma"){#vajab actuar paketti
  #ning parameetrite algvaartustamist
  fx = dlgamma(x_arg, ratelog = par[1], shapelog = par[2])
  qx = qlgamma(e_teor[,1], ratelog = par[1],
               shapelog = par[2])
}

df_fx = data.frame(x = x_arg/cSkaala, y = fx*cSkaala)
names(df_fx) = c("x_arg", "fx")

if(hist == TRUE){
  require(scales)
  #joonistame histogrammi andmetega tutvumiseks
  gHist = ggplot()#loome tausta
  gHist = gHist + theme_classic()
}

```

```

gHist = gHist + geom_histogram(aes(x = alvus/cSkaala
                                   , ..density..),
                               bins = tArv, alpha = .5,
                               color = "black",
                               fill = "white")
gHist = gHist + geom_line(data = df_fx,
                          aes(x = x_arg,
                              y = fx,
                              color = "red"),
                          size = 1)
gHist = gHist + theme(legend.position="none")
gHist = gHist + ylab("")
gHist = gHist + scale_y_continuous(labels = comma)
gHist = gHist + scale_x_continuous(labels = comma)
gHist = gHist + xlab("Kahjud")
gHist = gHist + ggtitle("Kahjud ja sobitatus jaotus")
gHist = gHist +
  theme(plot.title = element_text(hjust = 0.5))
#pealkiri keskele
}

if((kvant == TRUE) || (ekspKvant == TRUE)){
  #Loome andmestiku empiiriliste ja nende vastavate
  #teoreetiliste kvantiilidega
  qdf = data.frame(e_teor[,1], sort, qx)
  colnames(qdf) = c("alpha", "emp", "teor")

  if(kvant == TRUE){
    #teeme hajuvusdiagrammi
    gKvant = ggplot(qdf[qdf$alpha <= aLimKvant, ],
                   aes(x = teor/cSkaala, y = emp/cSkaala))
    #loome tausta
    gKvant = gKvant + theme_classic()
    gKvant = gKvant + geom_point(shape=1) #ringikesed
    #teeme seest tühjad
    gKvant = gKvant + ylab("Empiirilised kvantiilid")
    gKvant = gKvant + xlab("Teoreetilised kvantiilid")
    gKvant = gKvant + ggtitle("Kvantiil-kvantiil graafik")
    gKvant = gKvant +
      theme(plot.title = element_text(hjust = 0.5))
  }
}

```

```

#pealkiri keskele
gKvant = gKvant + scale_y_continuous(labels = comma)
gKvant = gKvant + scale_x_continuous(labels = comma)
gKvant = gKvant + geom_abline(slope = 1,
                             color = "red",
                             linetype = "solid",
                             size = 0.55)

}
if(ekspKvant == TRUE){

gEkspKvant = ggplot()
gEkspKvant = gEkspKvant + theme_classic()
gEkspKvant = gEkspKvant +
  geom_line(data = qdf[qdf$alpha <= aLimHist,],
            aes(x = alpha, y = teor/cSkaala,
                colour = "red"),
            linetype = "solid",
            size = 0.5)#teoreetilised kvantiilid
gEkspKvant = gEkspKvant +
  geom_line(data = qdf[qdf$alpha <= aLimHist,],
            aes(x = alpha, y = emp/cSkaala,
                colour = "green"), linetype = "solid",
            size = 0.5)#empiirilised kvantiilid
gEkspKvant = gEkspKvant +
  geom_line(data = ekspektiilid[
    ekspektiilid$alpha <= aLimHist,],
            aes(x = alpha, y = emp/cSkaala, colour = "red"),
            linetype = "dashed", size = 0.5)
#empiirilised ekspektiilid
gEkspKvant = gEkspKvant +
  geom_line(data = ekspektiilid[ekspektiilid$alpha <=
                                aLimHist,],
            aes(x = alpha, y = teor/cSkaala,
                colour = "green"), linetype = "dashed",
            size = 0.5)#teoreetilised ekspektiilid
gEkspKvant = gEkspKvant + xlab(expression(alpha))
gEkspKvant = gEkspKvant + ylab("")
gEkspKvant = gEkspKvant +
  ggtitle("Ekspektiilid ja kvantiilid")
}

```

```

    gEkspKvant = gEkspKvant +
        theme(plot.title = element_text(hjust = 0.5))
    #pealkiri keskele
    gEkspKvant = gEkspKvant + theme(legend.position="none")
    gEkspKvant = gEkspKvant +
        scale_y_continuous(labels = comma)
    gEkspKvant = gEkspKvant +
        scale_x_continuous(labels = comma)

    }
    }
}

if(hist == FALSE){
    if(kvant == FALSE){
        if(ekspKvant == FALSE){
            return(gEksp)
        }else{
            return(grid.arrange(gEksp, gEkspKvant, ncol = 2))
        }
    }else if(ekspKvant == FALSE){
        return(grid.arrange(gEksp, gKvant, ncol = 2))
    }else{
        return(grid.arrange(gEksp, gKvant, gEkspKvant, ncol = 2))
    }
}
else if(kvant == FALSE){
    if(ekspKvant == FALSE){
        return(grid.arrange(gEksp, gHist, ncol = 2))
    }else{
        return(grid.arrange(gEksp, gEkspKvant, gHist, ncol = 2))
    }
}
else if(ekspKvant == FALSE){
    return(grid.arrange(gEksp, gKvant, gHist, ncol = 2))
}
else{
    return(grid.arrange(gEksp, gKvant, gHist, gEkspKvant,
                        ncol = 2))
}
}
}

```



## **Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks**

Mina, Johannes Puusepp,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose „Ekspektil-ekspektil graafikud jaotuse sobivuse hindamiseks“, mille juhendaja on dotsent Meelis Käärik,
  - 1.1. reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;
  - 1.2. üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.
2. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.
3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartus, 09.05.2017